

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 18:08:00 ; Search time 115.49 Seconds
(without alignments)
4008.321 Million cell updates/sec

Title: US-09-593-828-3
Perfect score: 2044
Sequence: 1 ggctcgaggtccactgtgct.....aaaaaaaaaaaaaaaaaaaa 2044

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11328999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	400	19.6	2032	4 US-09-045-284A-1	Sequence 1, Appli
2	235.2	11.5	2156	2 US-08-899-514-1	Sequence 1, Appli
3	208.2	10.2	13865	3 US-09-009-217-11	Sequence 11, Appli
4	208.2	10.2	13865	3 US-09-009-656-11	Sequence 11, Appli
5	204.2	10.0	72928	3 US-09-009-913-1	Sequence 1, Appli
6	202.8	9.9	2713	2 US-08-916-901-6	Sequence 6, Appli
7	202.8	9.9	2713	4 US-09-154-602-6	Sequence 6, Appli
8	201.6	9.9	2115	1 US-08-395-800A-7	Sequence 7, Appli
9	201.2	9.8	22481	5 PCT-US95-07201-43	Sequence 43, Appli
10	200.2	9.8	6669	3 US-09-212-971-5	Sequence 5, Appli
11	200.2	9.8	6669	3 US-08-800-929A-5	Sequence 5, Appli
12	200.2	9.8	6669	4 US-09-617-053A-5	Sequence 5, Appli
13	198.4	9.7	282	1 US-08-133-629-8	Sequence 8, Appli
14	197.4	9.7	629	4 US-09-385-982-204	Sequence 204, App
15	196.8	9.6	2354	1 US-08-655-878-1	Sequence 1, Appli
16	196.8	9.6	20303	1 US-08-370-975B-6	Sequence 6, Appli
17	196.8	9.6	26764	1 US-08-370-975B-1	Sequence 1, Appli
18	196.6	9.6	12047	2 US-09-022-461-1	Sequence 1, Appli
19	196.4	9.6	1000	4 US-09-018-584A-32	Sequence 32, Appli
20	194.4	9.5	8353	3 US-08-611-587-1	Sequence 1, Appli
21	193.6	9.5	631	4 US-09-385-982-354	Sequence 354, App
22	193.4	9.5	4823	2 US-08-457-254-5	Sequence 5, Appli
23	193.4	9.5	4823	2 US-08-484-257-20	Sequence 20, Appli
24	193.4	9.5	4823	3 US-08-999-927-5	Sequence 5, Appli
25	193.4	9.5	4823	5 PCT-US94-08806-28	Sequence 28, Appli
26	193.4	9.5	4823	5 PCT-US95-01829-5	Sequence 5, Appli
27	193.4	9.5	4823	5 PCT-US95-16626-5	Sequence 5, Appli

28	193.4	9.5	11288	4 US-08-646-301A-1	Sequence 1, Appli
29	193.4	9.5	11288	4 US-08-481-968A-4	Sequence 4, Appli
30	193.2	9.5	246240	2 US-08-724-394A-20	Sequence 20, Appli
31	193.2	9.5	246240	2 US-08-724-394A-21	Sequence 21, Appli
32	193.2	9.5	246240	2 US-08-724-394A-22	Sequence 22, Appli
33	192.6	9.4	246240	2 US-08-724-394A-20	Sequence 20, Appli
34	192.6	9.4	246240	2 US-08-724-394A-21	Sequence 21, Appli
35	192.6	9.4	246240	2 US-08-724-394A-22	Sequence 22, Appli
36	191.6	9.4	36741	4 US-09-301-665-3	Sequence 3, Appli
37	191.4	9.4	1701	4 US-09-078-294-9	Sequence 9, Appli
38	191.2	9.4	28994	3 US-08-884-324-14	Sequence 14, Appli
39	191	9.3	14796	4 US-08-975-080-35	Sequence 35, Appli
40	191	9.3	14796	4 US-09-630-706-10	Sequence 10, Appli
41	191	9.3	35060	3 US-08-814-095-7	Sequence 7, Appli
42	189.4	9.3	152331	3 US-09-128-155-16	Sequence 16, Appli
43	189	9.2	4038	3 US-08-969-125-8	Sequence 8, Appli
44	189	9.2	4233	4 US-09-056-105-27	Sequence 27, Appli
45	188.4	9.2	1811	1 US-08-848-252-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-045-284A-1
; Sequence 1, Application US/09045284A
; Patent No. 6265192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107US1
; CURRENT APPLICATION NUMBER: US/09/045.284A
; CURRENT FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-045-284A-1

Query Match	19.6%;	Score 400;	DB 4;	Length 2032;
Best Local Similarity	63.6%;	Pred. No. 4.9e-69;		
Matches	645;	Conservative	0;	Mismatches 360;
			Indels	9;
			Gaps	2;
QY	335	gategtgtgcacgtgctggtgctgtccctcgtggtgcgtcggcgtcgcctcattcttcttgggccag	394	
Db	302	gagcgcgtgcacgtgctggtctgtcttcctggtcgtcgtcgtctgtcttcttcttcttcttggggcag	361	
QY	395	ctcttcagccagcaccgccgacgtctttctacctgatggagccgcgtggtggcatgtgtggacc	454	
Db	362	ctttttggcagcaccagatgtttctacctgatggagccgcgtggtggcatgtgtggatg	421	
QY	455	acctgtgcagggcagcgcgggcaacgctgcacatggccgtgcgcacctgatgcgctct	514	
Db	422	accttaagcagagcaccgcctggatgctgcacatggctgtgcgggatctgatacgggc	481	
QY	515	atcttttgcgacatggacgtgtttgatgcctacatgccacagag---ccgaaacctg	571	
Db	482	gtcttctgtgcacatgagcgtcttcttgatgcctacatggaaacctggtcccccggagacag	541	
QY	572	tcgccttttcaactgggcaacgagccgcgctgtgctgcgcgcgcctgcagcgc	631	
Db	542	tccagctctttcagtggtggagaacagccggccctgtgttctgcacctgcctgtgacatc	601	
QY	632	ttttcccgaggccaccatcagcaagcaggacgtatgcaagacactgtgcacgcggcagcca	691	
Db	602	atcccaaatgaaatcatcccccgggtcactgcaggtcctctgtgcagtaacagccc	661	
QY	692	ttcagcctggcccgaggcctgcctcctacagccacgtgtgtgtcgaagggtgcgc	751	

```

Db      662  tttgaggtggtggagaagccctgccgctcctacagccacgtggtgctcaaggaggtgcgc 721
QY      752  ttcttcaacctgcagggtgctctaccgctgctcagcgaccccgcgctcaaacctgcgcatac 811
Db      722  ttcttcaacctgcagtcacctctacccgctgctgaaagacccctccctcaaacctgcatac 781
QY      812  gtgcacctggtgcgcgaccccgccgctgctgcgctcccgaggcgcgccgcgcata 871
Db      782  gtgcacctggtgcgcgaccccgccgctggttccgttcccgagaacgcacaaaggagat 841
QY      872  ctggcacgcgacaaacgcatcgtgctggggcaccacaaacggcaagtgggtggagccgcacct 931
Db      842  ctcatgattgacagtcgcattgtgatggggcagcatgagcagaaactcaagaaggaggac 901
QY      932  cacctgcgcctgattcgcgaggtgtgcgcgacccacgtgcgcatcgccgaggccgcacaca 991
Db      902  caacctactatgtgatgcaggtcatctgcctgccaagccagctggagatctacaagaccatc 961
QY      992  ctcaagccgcaccccttctgcgcgcgcgctaccgcctggtgctgcttcgaggacctggcg 1051
Db      962  cagtccttgcccaaggccctgcaggaacgtacctgcttctgcgctatgaggacctggct 1021
QY      1052  cgggagccgctggcagagatccgcgcactctacgccttcaacggccctgacccctcacgcaca 1111
Db      1022  cgagccctgtgcccagacttcccgaaatgatgaattcgtgggattggaattcttgcgc 1081
QY      1112  cagctcgagccctggtatccacacatcacccacgggtcggggacgcggcaagccaatcgag 1171
Db      1082  catcttcagacctgggtgcataacatcacccgagggcaaggcaggtgta-----ccac 1135
QY      1172  gccttccatacttgcctaggaatgcgcgcaacgtctcccaggcctggcgccacgcgttg 1231
Db      1136  gctttccacacaaatgccaggatgcccttaattgtctcccaggcttgcgctggtctttg 1195
QY      1232  cccttcactaagatcctgcgcgtgcaggagggtgtgcgcgcgcgcgctgcagctgctgggc 1291
Db      1196  ccctatgaaaaggtttctcgaattcagaagaacgctgtgcgagtgccatgaatttgctgggc 1255
QY      1292  taccgacctgtgtactctgcggaccagcagcgtgacctacccctggatctggtg 1345
Db      1256  taccgcaogtcatgatctgaacaagaacagagaaacctgttgctggatcttctg 1309

RESULT 2
US-08-899-514-1
; Sequence 1, Application US/08899514
; Patent No. 5910581
; GENERAL INFORMATION:
;   APPLICANT: HABUCHI, OSAMI
;   APPLICANT: FUKUTA, MASAKAZU
;   TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
;   TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
;   TITLE OF INVENTION: FOR THE POLYPEPTIDE
;   NUMBER OF SEQUENCES: 9
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR, LLP
;   STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
;   CITY: NEWPORT BEACH
;   STATE: CALIFORNIA
;   COUNTRY: US
;   ZIP: 92660
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/899,514
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:

```



```

; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,901
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0367 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
;
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2713 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-916-901-6

```

Query Match	9.9%;	Score 202.8;	DB 2;	Length 2713;
Best Local Similarity	80.6%;	Pred. No. 7.1e-31;		
Matches 274;	Conservative 0;	Mismatches 62;	Indels 4;	Gaps 3;
QY 1664	actgctaataataataatgtgtgccaggccgggtgcggtggtcatgcctgtaatccc	1723		
Db				
QY 1086	ACTGTTTCATTGAAAAGAAATGGTGTCCGGCCAGGCGGGTGGCTCATGCCCTGTAATCCC	1145		
Db				
QY 1724	agcattttgagaggtgagcggggtggatcacctgaggtcagga-ttcaaaaaccagcctg	1782		
Db				
QY 1146	AGCACTTTGGGAGGCGGAGGCGGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCCTG	1205		
Db				
QY 1783	gccacaatagtgaaacccc-ctctctactactaaaaatgc-aaaaattagtcgggtggtg	1839		
Db				
QY 1206	GCCAAACATGGTGAACCCCTGTCTCTACTAAAAATACAAAAAATTAGCTGGCGCGGTG	1265		
Db				
QY 1840	gcacactcctgtaatcccagctacttaggaggtgaggtgggaaaaatcaactggactcca	1899		
Db				
QY 1266	GCGCATGCCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGATCACTGAACCCGG	1325		
Db				
QY 1900	aaggtggaggtgcagtaagctgaaatcatgccactgcaccctagcttgggtggcaaaagc	1959		
Db				
QY 1326	GAGGCAGAGTGTGCAGTGAGCCGAGATCATGCCATTGCACCTCCAGCCCTGGGTGACAGAGC	1385		
Db				
QY 1960	aaaactctatcaaaaaataataataaattgtttcaaaa	1999		
Db				
QY 1386	GAGACTCCATCTCAAAAAAATAATAATAAATAAATTATGAA	1425		
Db				

ADDRESSEE: Incyte Pharmaceuticals, Inc.
SIREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/154,602
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/916,901
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0367 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2713 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-154-602-6

	Query Match	9.9%;	Score 202.8;	DB 4;	Length 2713;
	Best Local Similarity	80.6%;	Pred. No. 7.1e-31;		
	Matches 274;	Conservative	0;	Mismatches 62;	Indels 4;
				Gaps	3;
QY	1664	actgctaaattaaattggtgccaggccgggtgcggtggtcattgctgaatccc	1723		
Db	1086	ACTGTTTCATTGAAAAAGAAATGGTGTCCGGCCAGGCGGGTGGCTCATGCCGTGAATCCC	1145		
QY	1724	agcaattttgagaggctgagcggttggtggtcacctgaggtcagga-ttcaaaaccagccgtg	1782		
Db	1146	AGCACTTTGGGAGGCGGAGGGGGGTGGATCACCTGAGGTTCAGGAGTTCGAGACCAGCCGTG	1205		
QY	1783	gccaacatagtgaacccc-ctctctactataaaaaatgc--aaaaattagtcgggcgtggtg	1839		
Db	1206	GCCAAACATGGTGAAACCCCTGTCTCTACTAAAAATACAAAAAATTAGTGGCGCGGTG	1265		
QY	1840	gcacactcctgtaatcccgactacttaggaggtgaggtgggaaaaatcacttggactcca	1899		
Db	1266	GCGCATGCCTGTATCCACAGCTACTCAGGAGGCTGAGGCAGGAGAAATCACTTGAACCCGG	1325		
QY	1900	aagggtgaggttgcaagtaagctgaaatcatgccactgcaccctagcttgggtggcaaacg	1959		
Db	1326	GAGGCAGAGGTTCAGCTGAGCCGAGATCATGCCATTGCACCTCCAGCCCTGGGTGACAGAGC	1385		
QY	1960	aaaactctatcaaaaaataataataaattgttcaaaa	1999		
Db	1386	GAGACTCCATCTCAAAAAAATAATAATAATAATATGAA	1425		

RESULT 7
US-09-154-602-6
; Sequence 6, Application US/09154602
; Patent No. 6300472
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, purvi
; TITLE OF INVENTION: RAB PROTEIN
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:


```

Db 9661 GCGCATGCCTGTAATCCAGCTACTTGGGAGGCTAAGGCAGGAGAATCACTTGAATCCA 9602
QY 1899 aaaggtgaggttgcaagctgaagtgaataatcatgccactgcaccctagcttggggtgcaaaag 1958
Db 9601 GGAAGTGGAGGTTGCAGTGAGCTGAGATCGGCCACTGCACICCCAGGCTGGGCAACAGAG 9542
QY 1959 caaaactctatcaaaaaataataataaa 1988
Db 9541 CAAGACTCCATCTCAAAAAAATAAAAAA 9512

RESULT 10
US-09-212-971-5/c
; Sequence 5, Application US/09212971B
; Patent No. 6107041
; GENERAL INFORMATION:
; APPLICANT: Korneluk, Robert G
; APPLICANT: Mackenzie, Alexander E
; APPLICANT: Liston, Peter
; APPLICANT: Baird, Stephen
; APPLICANT: Tsang, Benjamin K
; APPLICANT: Pratt, Christine
; TITLE OF INVENTION: DETECTION AND MODULATION OF IAPS AND
; TITLE OF INVENTION: NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
; TITLE OF INVENTION: DISEASE
; FILE REFERENCE: 07891/009002
; CURRENT APPLICATION NUMBER: US/09/212,971B
; CURRENT FILING DATE: 1998-12-16
; EARLIER APPLICATION NUMBER: 60/017,354
; EARLIER FILING DATE: 1996-04-26
; EARLIER APPLICATION NUMBER: 60/030,590
; EARLIER FILING DATE: 1996-11-14
; EARLIER APPLICATION NUMBER: 08/800,929
; EARLIER FILING DATE: 1997-02-13
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 6669
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variation
; LOCATION: (3677)...(3951)
; OTHER INFORMATION: n can be any nucleotide
US-09-212-971-5
```

```

Query Match          9.8%; Score 200.2; DB 3; Length 6669;
Best Local Similarity 79.6%; Pred. No. 2.8e-30;
Matches 249; Conservative 0; Mismatches 63; Indels 1; Gaps 1;

QY 1676 aattaaatgtgtgccaggccgggtgcggtgcctcatgcctgtaatcccagcattttgaga 1735
Db 1381 AAAAAACAGAGGCCTCTGCTGGTGCAGTGGCTCACGCTGTATCCAGCATTTTGGGA 1322
QY 1736 ggctgaggcgggtggatcacctgaggtcagga-ttcaaaaccagcctggccaaacatagtg 1794
Db 1321 GGCCAAGACAAAGTGGATCACTTGAGGTCAGGAGTTCAAACCAGCCTGGCCAAAATGGTG 1262
QY 1795 aaacccccctctactactaaaaatgcaaaaattagtcggcggtggtggcacactcctgtaat 1854
Db 1261 AAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTACCGCGCCTGTAAT 1202
QY 1855 cccagctacttagggctgaggtgggaaaaatcacttggactccaaaggtgaggttgca 1914
Db 1201 CCCAGCTACTCCGGAGGCTGAGGCAGGAGAAATCAATTTGAGCCTGGGAGGTGGAGTTGCA 1142
QY 1915 gtaagctgaataatcatgccactgcaccctagcttggtggcaaaagcaaaaacttatcaaaa 1974
Db 1141 GTGGGCGGAAATAGTCCCACTACACTCCAGCCTGGGTGACACAGCAAGACTCTGTCTCAA 1082
QY 1975 aaataataataa 1987
```

```

Db 1081 AAATAAAATAAAAA 1069

RESULT 11
US-08-800-929A-5/c
; Sequence 5, Application US/08800929A
; Patent No. 6133437
; GENERAL INFORMATION:
; APPLICANT: Korneluk, Robert G
; APPLICANT: Mackenzie, Alexander E
; APPLICANT: Liston, Peter
; APPLICANT: Baird, Stephen
; APPLICANT: Tsang, Benjamin K
; APPLICANT: Pratt, Christine
; TITLE OF INVENTION: DETECTION AND MODULATION OF
; TITLE OF INVENTION: IAPS AND NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERAT
; TITLE OF INVENTION: DISEASE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,929A
; FILING DATE: 13-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/030,590
; FILING DATE: 14-NOV-1996
; APPLICATION NUMBER: 60/017,354
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bieker-Brady, Kristina
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: 07891/009001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6669 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-800-929A-5

Query Match          9.8%; Score 200.2; DB 3; Length 6669;
Best Local Similarity 79.6%; Pred. No. 2.8e-30;
Matches 249; Conservative 0; Mismatches 63; Indels 1; Gaps 1;

QY 1676 aattaaatgtgtgccaggccgggtgcggtgcctcatgcctgtaatcccagcattttgaga 1735
Db 1381 AAAAAACAGAGGCCTCTGCTGGTGCAGTGGCTCACGCTGTATCCAGCATTTTGGGA 1322
QY 1736 ggctgaggcgggtggatcacctgaggtcagga-ttcaaaaccagcctggccaaacatagtg 1794
Db 1321 GGCCAAGACAAAGTGGATCACTTGAGGTCAGGAGTTCAAACCAGCCTGGCCAAAATGGTG 1262
QY 1795 aaacccccctctactactaaaaatgcaaaaattagtcggcggtggtggcacactcctgtaat 1854
Db 1261 AAACCCCAICTCTACTAAAAAATACAAAAATTAGCTGGGTGTGGTGGTACCGCGCCTGTAAT 1202
```

QY 1855 cccagctacttaggaggtgggaaatcacttggactccaaagggtggaggttgca 1914
|||||
Db 1201 CCCAGCTACTCCGGAGGCTGAGCAGGAGAAATCATTTGAGCCTGGGAGGTGGAGGTTGCA 1142
QY 1915 gtaagctgaaatcatgccactgcaaccctagcttgggtggcaagcaaaactctatcaaaa 1974
|||
Db 1141 GTGGGCGGAAATAGTGCCACTACACTCCAGCCTGGGTGACACAGCAAGACTCTGTCTCAA 1082
QY 1975 aaataattaataa 1987
|||||
Db 1081 AAATAAAATAAAAA 1069

RESULT 12
US-09-617-053A-5/C
; Sequence 5, Application US/09617053A
; Patent No. 6300492
; GENERAL INFORMATION:
; APPLICANT: Korneluk, Robert G
; APPLICANT: Mackenzie, Alexander E
; APPLICANT: Liston, Peter
; APPLICANT: Baird, Stephen
; APPLICANT: Tsang, Benjamin K
; APPLICANT: Pratt, Christine
; TITLE OF INVENTION: DETECTION AND MODULATION OF IAPS AND
; TITLE OF INVENTION: NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
; TITLE OF INVENTION: DISEASE
; FILE REFERENCE: 07891/009003
; CURRENT APPLICATION NUMBER: US/09/617,053A
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/800,929
; PRIOR FILING DATE: 1997-02-13
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 6669
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variation
; LOCATION: (3677)...(3951)
; OTHER INFORMATION: n can be any nucleotide
US-09-617-053A-5

Query Match 9.8%; Score 200.2; DB 4; Length 6669;
Best Local Similarity 79.6%; Pred. No. 2.8e-30;
Matches 249; Conservative 0; Mismatches 63; Indels 1; Gaps 1;
QY 1676 aattaaatgtgtgccaggccgggtgggtggctcatgcctgttaatccacagcattttgaga 1735
|||
Db 1381 AAAAAACAGAGGCCCTCTGCCTGGTGCACTGGCTCACGCCCTGTAAATCCACAGCAITTTGGGA 1322
QY 1736 ggctgaggcgggtggatcacctgaggtcagga-ttcaaaaccagcctgggccaaacatagt 1794
|||
Db 1321 GGCCAAGACAAGTGGATCATTGAGGTGAGGAGTTCAAACCAGCCTGGCCAAATGGTG 1262
QY 1795 aaacccctctctactactaaaaatgcaaaaattagtcggcggtggtggcacactcctgtaat 1854
|||||
Db 1261 AAACCCCATCTCTACTAAAAATACAAAATTAGCTGGGTGGTGTTAGCCTGGGAGGTTGCA 1202
QY 1855 cccagctacttaggaggtgagtggtgggaaatcaacttggactccaaagggtggaggttgca 1914
|||||
Db 1201 CCCAGCTACTCCGAGGCTGAGGCAGGAGAAATCATTTAGCCTGGGAGGTTGCA 1142
QY 1915 gtaagctgaaatcatgccactgcaaccctagcttgggtggcaagcaaaactctatcaaaa 1974
|||
Db 1141 GTGGGCGGAAATAGTCCACTACACTCCAGCCTGGGTGACACAGCAAGACTCTGTCTCAA 1082
QY 1975 aaataattaataa 1987
|||||
Db 1081 AAATAAAATAAAAA 1069

RESULT 13
US-08-133-629-8
; Sequence 8, Application US/08133629
; Patent No. 5597694
; GENERAL INFORMATION:
; APPLICANT: Munroe, David J.
; APPLICANT: Housman, David E.
; TITLE OF INVENTION: AMPLIFICATION OF NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/133,629
; FILING DATE: 07-OCT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Greer, Helen
; REGISTRATION NUMBER: 36,816
; REFERENCE/DOCKET NUMBER: M0828/7001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX: 92-1742 EZEKIEL
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 282 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-133-629-8

Query Match 9.7%; Score 198.4; DB 1; Length 282;
Best Local Similarity 82.5%; Pred. No. 3.1e-30;
Matches 231; Conservative 6; Mismatches 42; Indels 1; Gaps 1;
QY 1692 gcccgggtgcggtgggtcctcatgcctgttaatccacagcattttgagaggtgagcggtgga 1751
|||
Db 1 GGCTGGGCGTGGTGGCTCACACCTGTAAATCCACAGCACTTTGGGAGGCCGAGGTGGGTTGGA 60
QY 1752 tcacctgaggtcagga-ttcaaaaccagcctggccaacatagtgaaacccctctctact 1810
|||
Db 61 TCACCTGAGGTCAGGAGTCAAGACCAGCCTGGCAACATGGTGAACCCGCTCTCTACT 120
QY 1811 aaaaatgcaaaaattagtcggcggtggtggcacactcctgttaatccacagctacttaggag 1870
|||
Db 121 AAAAAACAAAAATTAGCCGGCGCTGGTGGCGCGCCTGTAAICCCAGCTACTCGGGAG 180
QY 1871 gctgaggtgggaaatcaacttggactccaaagggtggaggttgagtaagctgaaatcatg 1930
|||
Db 181 GCTGAGGCAGGAGAAATCGCTTGAAACCCAGGAGGTGAGGTYTCAGTGGCCGWTGTCGG 240
QY 1931 ccaactgcaccctagcttgggtggcaagcaaaactctatc 1970
|||
Db 241 CCACTGCACCTCCAGCCTGGGCAACAGAGYRAGACTCYRTC 280

RESULT 14
US-09-385-982-204/C
; Sequence 204, Application US/09385982
; Patent No. 6262334

; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS: II
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-08-30
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 204
; LENGTH: 629
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(629)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-204

Query Match 9.7%; Score 197.4; DB 4; Length 629;
Best Local Similarity 78.8%; Pred. No. 5.7e-30;
Matches 234; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 1692 ggcgggtgcgggtgcctgcctgtaatacccagcatttgagaggtgaggggtgga 1751
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
309 GCCAGGCGGCGGCTCATGCTGTTATCCCTAGCACTCTGGNAGGCTAAGCGGGTGA 250
QY 1752 tcacctgaggtcaggattcaaaaccagcctggcccaacatagtgaacccctcttacta 1811
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
249 TCACCTGAGCTCAGAAGTTCAAGACCAGCTGGCCACACATATGAATCCCATCTNTACTA 190
QY 1812 aaatgcaaaattagtcggcgtggtggcacactcctgttaatacccagctacttaggag 1871
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
189 AAATACAAAAATAGTCAGGCGTGGTGACATGCCTGTAGTCCAGCTACTCGGGAGG 130
QY 1872 ctgaggtgggaaatcacttggaactcaaaaggtggaggttcagtaagctgaaatcatgc 1931
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
129 TTGAGCGCAGGAGATCACTTGAAACCGGGGAGCGGAGGTTGTCAGTGAGCCGATGTCGTGC 70
QY 1932 cactgcaccctagcttgggtggcaagcaaaactctatcaaaaaataataataaa 1988
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
69 CACTGCACTCTAGCCTGGGTGACAGACTGAGACTCCGTCTCAAAAAAATAAAAAA 13

RESULT 15
US-08-655-878-1
; Sequence 1, Application US/08655878
; Patent No. 5827713
; GENERAL INFORMATION:
; APPLICANT: FUKUIA, MASAKAZU
; APPLICANT: HABUCHI, OSAMI
; TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:
; STREET:
; CITY:
; STATE:
; COUNTRY:
; ZIP:
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/655,878
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:
; TELEFAX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2354
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Chick
; TISSUE TYPE: Embryo chondrocyte
; FEATURE:
; NAME/KEY: CDS
; FEATURE: LOCATION: 211..1584
; FEATURE: IDENTIFICATION METHOD: P
; FEATURE:
; NAME/KEY: sig_peptide
; FEATURE: LOCATION: 211..309
; FEATURE: IDENTIFICATION METHOD: P
; FEATURE:
; NAME/KEY: mat_peptide
; FEATURE: LOCATION: 310..1584
; FEATURE: IDENTIFICATION METHOD: P
; FEATURE:
; NAME/KEY: transmembrane domain
; FEATURE: LOCATION: 280..321
; FEATURE: IDENTIFICATION METHOD: P
; FEATURE:
; NAME/KEY: potential N-glycosylation site
; FEATURE: LOCATION: 394..402
; FEATURE: IDENTIFICATION METHOD: S
; FEATURE:
; NAME/KEY: potential N-glycosylation site
; FEATURE: LOCATION: 427..435
; FEATURE: IDENTIFICATION METHOD: S
; FEATURE:
; NAME/KEY: potential N-glycosylation site
; FEATURE: LOCATION: 493..501
; FEATURE: IDENTIFICATION METHOD: S
; FEATURE:
; NAME/KEY: potential N-glycosylation site
; FEATURE: LOCATION: 916..924
; FEATURE: IDENTIFICATION METHOD: S
; FEATURE:
; NAME/KEY: potential N-glycosylation site
; FEATURE: LOCATION: 1405..1413
; FEATURE: IDENTIFICATION METHOD: S
; FEATURE:
; NAME/KEY: potential N-glycosylation site
; FEATURE: LOCATION: 1537..1545
; FEATURE: IDENTIFICATION METHOD: S
US-08-655-878-1

Query Match 9.6%; Score 196.8; DB 1; Length 2354;
Best Local Similarity 52.4%; Pred. No. 9.9e-30;
Matches 575; Conservative 0; Mismatches 487; Indels 36; Gaps 5;

QY 269 cagaccacctgcctcctgctcttcatcatctcccgccagggccctccatccccagccggc 328
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
475 CTGAGGACGCGGCTCGGCAACGTCACCTTGCACTGGCTGGGAGAGCTGGGCATAGCAGCC 534

us-09-593-828-3.rni

Fri Feb 1 20:20:51 2002

search completed: January 31, 2002, 18:09:45
Job time: 10729 sec

QY	329	ggc	gag	gat	cgt	gtg	ca	cgt	gtc	gtg	gtc	gtg	gc	gtc	ggg	ctc	atc	ttc	ttg	388
Db	535	CCAGAG	CGCGGG	GCGAT	GTCC	TGTA	TGGC	CA	CCAC	ACGCA	CGCAC	CGGCT	CCCT	CTCG	TT					594
QY	389	ggc	cag	ctt	cag	ccag	ca	cccc	gac	gtc	tctt	tac	ctg	atg	gag	ccc	gct	ggc	atg	448
Db	595	GGG	AGT	TTCT	TA	ACC	AGC	AGG	CA	ACAT	ATTCT	TA	CTCT	TTG	AG	CCCC	CTGT	GG	CA	654
QY	449	tgg	acc	acc	ctg	tc	-----	gc	agg	gc	agc	ggc	aa	cgt	gc	ac	at	ggc	gctg	496
Db	655	GAG	AG	ACGG	TCA	CTTT	TG	AG	CC	AGGG	GGGCC	CA	AC	CGG	TGG	GCT	CG	CCCT	GGT	714
QY	497	cgc	gac	ctg	atg	cgc	ctc	tat	ctt	ttt	gtg	cg	aca	tg	gac	gtg	ttt	gat	gc	556
Db	715	CGC	AG	CTG	CTGC	AG	CTC	CTC	CT	CG	AC	CTCT	AC	ATTCT	GG	AG	AGCT	TTCA	TCTCA	774
QY	557	cag	agc	gaa	aac	ctg	tc	-----	gc	ctt	ttt	caa	act	ggg	ca	ac	gag	cgc	gcg	607
Db	775	CCAG	CGCC	GAG	AG	CA	ACCT	AA	CTG	TGCC	CTGT	CC	GG	CGGG	GCTC	CAG	CC	ACT	CA	834
QY	608	tg	ctc	gc	gc	cccc	gc	ctg	ca	gc	gc	ctt	tcccc	gag	gc	ac	ca	tcag	ca	667
Db	835	TGT	GAG	GAG	CCG	CTCT	GC	AC	CC	AG	CTCA	AG	AG	GTCT	TT	G	AG	AGT	ACC	894
QY	668	aag	ac	act	gtg	ca	cgc	ggc	ag	cc	attc	ag	cc	tg	gc	ccc	ggg	ag	gc	727
Db	895	A	CCG	CCG	CTG	CGG	CCCT	CT	CA	AC	ATC	AC	CGT	GG	CAG	CTGA	AG	CA	TGCC	954
QY	728	ca	cgt	ggt	gtc	aa														

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 18:35:34 ; Search time 4771.45 Seconds
(without alignments)
4055.621 Million cell updates/sec

Title: US-09-593-828-4
Perfect score: 1173
Sequence: 1 atgtggctgccacggttctc.....gttgggcacgcctgactga 1173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_om.*
- 20: em_or.*
- 21: em_ov.*
- 22: em_pat.*
- 23: em_ph.*
- 24: em_pl.*
- 25: em_ro.*
- 26: em_sts.*
- 27: em_sy.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htgo_hum.*
- 31: em_htgo_inv.*
- 32: em_htgo_rod.*
- 33: em_htg_hum.*
- 34: em_htg_inv.*
- 35: em_htg_rod.*
- 36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1173	100.0	1462	9	AF176839	AF176839 Homo sapi
2	1173	100.0	2170	9	AF176838	AF176838 Homo sapi
3	1173	100.0	3278	9	AF246718	AF246718 Homo sapi
4	1173	100.0	71503	9	AF219991	AF219991 Homo sapi
5	1173	100.0	159072	2	AC026419	AC026419 Homo sapi
6	1173	100.0	181566	2	AC025287	AC025287 Homo sapi
c	1053.8	89.8	179065	2	AC009163	AC009163 Homo sapi
8	924.6	78.8	2544	9	AF219990	AF219990 Homo sapi
9	924.6	78.8	3786	9	AF280086	AF280086 Homo sapi
c	924.6	78.8	159072	2	AC026419	AC026419 Homo sapi
11	924.6	78.8	208185	2	AC009105	AC009105 Homo sapi
12	745.2	63.5	1740	10	AF176841	AF176841 Mus muscu
13	745.2	63.5	1989	10	AF176840	AF176840 Mus muscu
14	400	34.1	2032	9	AF131235	AF131235 Homo sapi
15	398.4	34.0	1333	9	AF149783	AF149783 Homo sapi
16	398.4	34.0	1992	9	AF280088	AF280088 Homo sapi
17	398.4	34.0	2011	9	AK026635	AK026635 Homo sapi
18	398.4	34.0	183133	2	AC010547	AC010547 Homo sapi
19	358.8	30.6	2201	10	AF109155	AF109155 Mus muscu
20	354	30.2	1926	10	AF131236	AF131236 Mus muscu
c	351.6	30.0	179065	2	AC009163	AC009163 Homo sapi
22	235.2	20.1	2156	6	AR071396	AR071396 Sequence
23	235.2	20.1	2156	6	E14937	E14937 Human mRNA
24	235.2	20.1	2731	9	AB012192	AB012192 Homo sapi
25	235.2	20.1	6961	9	AB017915	AB017915 Homo sapi
c	235.2	20.1	182677	2	AC073370	AC073370 Homo sapi
27	235.2	20.1	196465	9	AC022392	AC022392 Homo sapi
28	196.8	16.8	2354	5	CHKC6ST	D49915 Chicken mRN
29	196.8	16.8	2354	6	AR050577	AR050577 Sequence
30	196.8	16.8	2354	6	E12369	E12369 DNA encodin
c	196.2	16.7	137499	2	AC015931	AC015931 Homo sapi
32	196.2	16.7	165762	2	AP001582	AP001582 Homo sapi
c	196.2	16.7	174187	2	AP002792	AP002792 Homo sapi
34	196.2	16.7	215647	2	AC068591	AC068591 Homo sapi
35	190.6	16.2	1425	10	AF178689	AF178689 Rattus no
36	181	15.4	1654	10	AB008937	AB008937 Mus muscu
37	153.8	13.1	2227	9	AF083066	AF083066 Homo sapi
38	153.8	13.1	3029	9	AK023268	AK023268 Homo sapi
39	153.8	13.1	3535	9	AB021124	AB021124 Homo sapi
40	153.8	13.1	4862	9	AB021125	AB021125 Homo sapi
c	153.8	13.1	169753	2	AC055737	AC055737 Homo sapi
42	150.8	12.9	182575	2	AC018450	AC018450 Homo sapi
43	148.8	12.7	2393	9	AB014679	AB014679 Homo sapi
44	148.8	12.7	2393	9	AB014680	AB014680 Homo sapi
45	148.8	12.7	2409	6	E28527	E28527 Polypeptide

ALIGNMENTS

RESULT 1

AF176839

LOCUS

DEFINITION

AF176839 1462 bp DNA
Homo sapiens intestine N-acetylglucosamine 6-O-sulfotransferase
(I-GlcNAc-6-ST) gene, complete cds.
PRI 22-SEP-1999

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1462)

Lee, J.K., Bhakta, S., Rosen, S.D. and Hemmerich, S.

Cloning and characterization of a mammalian

N-acetylglucosamine-6-sulfotransferase that is highly restricted to

intestinal tissue

Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)

99423499

2 (bases 1 to 1462)

ROSEN, S. D. and HEMMERICH, S.

AUTHORS Lee, J.-K., Bhakta, S., Rosen, S.D. and Hemmerich, S.
TITLE Direct Submission
JOURNAL Submitted (10-AUG-1999) Respiratory Diseases, Roche Bioscience,
3601 Milliken Avenue, Palo Alto, CA 94304, USA

EXAMPLES	Location/Qualifiers
1. The first two	
2. The first two	
3. The first two	
4. The first two	
5. The first two	
6. The first two	
7. The first two	
8. The first two	
9. The first two	
10. The first two	
11. The first two	
12. The first two	
13. The first two	
14. The first two	
15. The first two	
16. The first two	
17. The first two	
18. The first two	
19. The first two	
20. The first two	
21. The first two	
22. The first two	
23. The first two	
24. The first two	
25. The first two	
26. The first two	
27. The first two	
28. The first two	
29. The first two	
30. The first two	
31. The first two	
32. The first two	
33. The first two	
34. The first two	
35. The first two	
36. The first two	
37. The first two	
38. The first two	
39. The first two	
40. The first two	
41. The first two	
42. The first two	
43. The first two	
44. The first two	
45. The first two	
46. The first two	
47. The first two	
48. The first two	
49. The first two	
50. The first two	
51. The first two	
52. The first two	
53. The first two	
54. The first two	
55. The first two	
56. The first two	
57. The first two	
58. The first two	
59. The first two	
60. The first two	
61. The first two	
62. The first two	
63. The first two	
64. The first two	
65. The first two	
66. The first two	
67. The first two	
68. The first two	
69. The first two	
70. The first two	
71. The first two	
72. The first two	
73. The first two	
74. The first two	
75. The first two	
76. The first two	
77. The first two	
78. The first two	
79. The first two	
80. The first two	
81. The first two	
82. The first two	
83. The first two	
84. The first two	
85. The first two	
86. The first two	
87. The first two	
88. The first two	
89. The first two	
90. The first two	
91. The first two	
92. The first two	
93. The first two	
94. The first two	
95. The first two	
96. The first two	
97. The first two	
98. The first two	
99. The first two	
100. The first two	

```

1. .1462
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q23.1"
/tissue_type="intestine"
<229. .>1401
/gene="I-GlcNAc-6-ST"
/product="N-acetylglucosamine 6-O-sulfotransferase"
<229. .>1401
/gene="I-GlcNAc-6-ST"
229. .1401
/gene="I-GlcNAc-6-ST"
/note="possibly involved in biosynthesis of mucosal
ligands for T-selectin"

```

```

/codon_start=1
/product="N-acetylglucosamine 6-O-sulfotransferase"
/protein_id="AAD56001.1"
/db_xref="GI:5917708"
/translation="MWLPFRFSKTVTVLLAQTTCLLFTITSRPGPSPAGGEDRVHV
LVLSRRSGSFLGQLFSQHPDVFYLMPEAHVWTTLSQGSAAATHMAVRDLMRSIFL
CDMDVFAYMQSRNLSAFFENWATSRALCSPPACSAFPRGTISKODVCKTLCTRQPF
LAREACRSYSHVLKEVREFNLQVLYPLSDPALNLRIVHLVRDPAVLRSREAAGPI
LARDNGIVLTNGKWEADPHLRIREVCRSHVRIAEATLKPPFLRIGRYLRVREFD
LAREPLAEIRALYAFGLTLTPQLEAWIHNTHGSGIGKPIEAFTFSRNARNVSOAW
RHALPTKILRVQEVCAQALQLLGYRPPVSADQQQRDLTLDLVLPGRPDHFSWASPD"
239 a 510 c 446 g 267 t

```

Db	709	CTGGCCCGGAGGCCCTGCCGCTCCTACAGCCACCGTGGTGGTCTCAAGGAGGTGCGCTTCTTC	768
QY	541	aacctgcaggtgctctaccogctgctcagcgacccccgcgtcaacctgcgcatcggtgcaac	600
Db	769	AACCTGCAGGTGCTCTACCCGCTGCTCAGCGACCCCGCGCTCAACCTGCGCATCGTGCAC	828
QY	601	ctggtgcgcgacccgcggcgctgtcgctccccgggagggcgggcccgatactggca	660
Db	829	CTGGTGGCGACCCGCGGCGCGTGGCTCCGGAGCGGGCGGCCGATACCTGGCA	888
QY	661	cgcacaacggcatcgtgctggggcaccacaggcaagtgggtggaggccgacccctcacctg	720
Db	889	CGCGACAACGGCATCGTGTGGGCACCAACAGGCAAGTGGTGGAGGCGCCCTACCTG	948
QY	721	cgctgattcgcgaggtgtgcgcgcagccacgtgcgcatcgccgagggccgccacactcaag	780
Db	949	CGCCTGATTCGCGAGGTGTGCCGCAGCCACGTGGCATCGCCGAGGCGCCACACTCAAG	1008
QY	781	ccgcacaccttcctgcgcggccgctaccgcctggtgagcttcgaggaacctggcgcgag	840
Db	1009	CCGCCACCCCTTCCTGCGGCGCGCTACCGCCCTGGTGGCTTCGAGGACCTGGCGGGGAG	1068
QY	841	ccgctggcagagatccgcgcactctacgccttcacoggcctgacctcacgccacagctc	900
Db	1069	CCGCTGGCAGAGATCCGCGCACTCTACGCCCTTCACCGGCCCTGACCCCTCACGCCACAGCTC	1128
QY	901	gaggcctggatccacaacatcacccacgggtcggggacggcaagccaatcgaggccttc	960
Db	1129	GAGGCCIGGATCCACAACATACCCACGGGTGCGGGATCGSSCAAGCCAAATCGAGGCCTTC	1188
QY	961	catactcgtctaggaatgcggcaacgtctctccaggcctggcgccacgcgttgcccttc	1020
Db	1189	CATACTTCGTCTAGGAATGCGGCAACGTCTCCCAGGCCCTGGCGCCACGCGTGGCCCTTC	1248
QY	1021	actaagatcctgcgcgtgcaggaggtgtgcgcgcgcgctgcagctgctgggctaccgg	1080
Db	1249	ACTAAGATCCTGCGGTGCAGGAGGTGTGCGCGCGCGCGCTGCAGCTGCTGGGTACCGG	1308
QY	1081	cctgttactctgcggaccagcagcgtgacctcacccctggatctggtgctgccacgagcc	1140
Db	1309	CCTGTGTACTCTGCGGACCAGCAGCGGTGACCTACCCCTGGATCTGGTGGCTGCCACGAGC	1368
QY	1141	ccagaccacttcagctgggcatcgccctgactga	1173

RESULT

				PRI	22-SEP-1999
AF176838				mRNA	
LOCUS	AF176838	2170 bp			
DEFINITION	Homo sapiens intestine N-acetylglucosamine 6-O-sulfotransferase				
	(X-CLONAC-6-ST) mRNA, complete cds.				

481 ctggcccgagggcctgccgctcctacagccacgtggtgctcaaggagggtgcgcttcttc 540


```
repeat_region /rpt_type=dispersed
complement(462..771)
/rpt_family="Alu"
repeat_region /rpt_type=dispersed
772..945
/rpt_family="L2"
repeat_region /rpt_type=dispersed
1169..1338
/rpt_family="MIR"
/rpt_type=dispersed
complement(1553..1804)
/rpt_family="Alu"
repeat_region /rpt_type=dispersed
2616..2916
/rpt_family="Alu"
/rpt_type=dispersed
3309..3532
/note="similar to EST H12297 (EST cluster Hs.31147)"
3865..4028
/note="similar to EST H12297 (EST cluster Hs.31147)"
4099..4406
/rpt_family="Alu"
/rpt_type=dispersed
4793..5250
/rpt_family="L2"
/rpt_type=dispersed
complement(5251..5329)
/rpt_family="MIR"
/rpt_type=dispersed
complement(5768..6080)
/rpt_family="Alu"
/rpt_type=dispersed
complement(6081..6265)
/rpt_family="Alu"
6111..6567
/note="similar to EST AA001593 (EST cluster Hs.6853)"
8035..8350
/rpt_family="Alu"
/rpt_type=dispersed
10545..11049
/note="similar to EST AI950923 (EST cluster Hs.6853)"
11043..11195
/rpt_family="Alu"
/rpt_type=dispersed
11200..11222
/rpt_family="Simple_repeat"
/rpt_type=tandem
11570..11871
/rpt_family="Alu"
/rpt_type=dispersed
complement(12065..12198)
/rpt_family="L1"
/rpt_type=dispersed
complement(12205..12493)
/rpt_family="Alu"
/rpt_type=dispersed
complement(12813..13110)
/rpt_family="Alu"
/rpt_type=dispersed
join(14185..14228,16169..16263,17712..20821)
/gene="CHST5"
/product="intestinal N-acetylglucosamine
6-O-sulfotransferase"
14185..20821
/gene="CHST5"
repeat_region /rpt_family="Simple_repeat"
14681..14713
/rpt_type=tandem
complement(14715..14990)
/rpt_family="Alu"
/rpt_type=dispersed
complement(14995..15288)
```

```
repeat_region /rpt_family="Alu"
/rpt_type=dispersed
complement(15292..15589)
/rpt_family="Alu"
repeat_region /rpt_type=dispersed
15668..15718
/rpt_family="Alu"
repeat_region /rpt_type=dispersed
15977..16098
/rpt_family="MIR"
/rpt_type=dispersed
complement(17298..17433)
/rpt_family="MIR"
/rpt_type=dispersed
complement(18001..18310)
/rpt_family="Alu"
/rpt_type=dispersed
complement(18420..18591)
/rpt_family="MIR"
/rpt_type=dispersed
18732..18767
/rpt_family="Simple_repeat"
/rpt_type=tandem
18967..20202
/gene="CHST5"
/note="I-GlcNAc6ST; carbohydrate sulfotransferase 5;
expressed in the small intestine and colon"
/codon_start=1
/product="intestinal
N-acetylglucosamine-6-O-sulfotransferase"
/protein_id="AAG26326.1"
/db_xref="GI:11023148"
/translation="MGMRAVPKVAHSTRPPAARMWLPFSKTVTVLLAQTCLL
LFIISRPGSPSPAGGEDRVHVLVLSWRSGSFLGQFSDVDVFLMEPAWHVWTTL
SQGSAATLHMAVRDLMSIFLCDMDVDAYMPQSRNLSAFNWATSALCSPACSAF
PRGTISKQDVCKILCTRQPFSLAREACRSYSHVVLKEVRFNLOVLYPLSDPALNR
TVHLVRDPRAVLRSREAAGPILARDNGIVLGINGWVEADPHRLIREVCRSHVRIAE
AATLKPPFLRGYLVRFEDLAREPLAEIRALYAFTGLTLPQLEAWIHNIHCSGI
GKPIEAFHTSSRNARNVSOAWRHALPTKILRVQEVCAQALQLGYRPVYSADQQRDL
TLDLVLPGRPDHFSWASPD"
20504..20806
/rpt_family="Alu"
/rpt_type=dispersed
20869..20961
/rpt_family="L1"
/rpt_type=dispersed
20962..21200
/rpt_family="MER4-group"
/rpt_type=dispersed
21325..21429
/rpt_family="MER4-group"
/rpt_type=dispersed
21430..21722
/rpt_family="Alu"
/rpt_type=dispersed
21725..22007
/rpt_family="Alu"
/rpt_type=dispersed
22008..22040
/rpt_family="Simple_repeat"
/rpt_type=tandem
22041..22313
/rpt_family="MER4-group"
/rpt_type=dispersed
22325..22593
/rpt_family="Alu"
/rpt_type=dispersed
22594..22949
/rpt_family="MER4-group"
/rpt_type=dispersed
23116..23221
/rpt_family="Simple_repeat"
/rpt_type=tandem
```

```

Query Match          100.0%;   Score 1173;   DB 9;   Length 71503;
Best Local Similarity 100.0%;   Pred. No. 7.7e-162;
Matches 1173;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```

	721	cgcctgattcgcgaggtgtgccgcaggccacgtcgcatcgccgagggccgccacctcaag	780
DY			
Db	19750	CGCCTGATTGCGAGGTGTGCCGCAGCCACGTCGGCATCGCGGAGCGCCACACTCAAG	19809
QY	781	ccgccacccttctcgcgcgcgtaccgcctggtgcgcttcgaggaactggcgcgag	840
Db	19810	CGCCACCCTTCTTGCGCGGCGCTACCGCCTGGTGGCTTCGAGGACCTGGCGGGAG	19869
QY	841	ccgctggcagatccgcgcacttaagccttacccggcctgacccctcacgccacagctc	900
Db	19870	CCGCTGGCAGAGATCCGGCGCACTTACGCCCTTACCGGCTGACCTCACGCCACAGCTC	19929
QY	901	gaggcctggatccacaacatccccacgggtcggggatcggaagccaatogagccctc	960
Db	19930	GAGGCCTTGGATCCACAACATCACCCACGGTCTCCAGGCTCGGGGATCGGCAAGCCAATCGAGGCCTTC	19989
QY	961	catacttgttagaatgcgcgaacgtctcccaggcctggcgaacgcttgcccttc	1020
Db	19990	CATACTTGCTTAGGAATGCGGCAACGTCCTCCAGGCTGGCGCCACCGTGGCCCTTC	20049
QY	1021	actaagatcctgcgctgcaggagtgtgcgcgcgcgctgcagctgctgggtaccgg	1080
Db	20050	ACTAAGATCCTGCGCGTGCAGGAGGTGTGCGCGCGCGCTGCAGCTGCTGGGTACCGG	20109
QY	1081	cctgtgtactctgcgaaccagcgtgacctcaccctggatctggtgctgccagaggc	1140
Db	20110	CCtGTGTACTCTGCGGACCAGCAGCGTGACCTCACCCCTGGATCTGGTGTGCCACGAGC	20169
QY	1141	ccagaccacttcagctgggcatcgctgactga	1173
Db	20170	CCAGACCACITCAGCTGGGATCGCCTGACTGA	20202
RESULT	5		
AC026419	AC026419	159072 bp	DNA
LOCUS	Homo sapiens chromosome 5 clone CTD-2113H21,	HTG	20-APR-2001
DEFINITION	SEQUENCE, 16 unordered pieces.		WORKING DRAFT
ACCESSION	AC026419		
VERSION	AC026419.3	GI:13699674	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 159072)		
AUTHORS	DOE Joint Genome Institute.		
TITLE	Sequencing of Human Chromosome 5		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 159072)		
AUTHORS	DOE Joint Genome Institute.		
TITLE	Direct Submission		
JOURNAL	Submitted (22-MAR-2000) Production Sequencing Facility, DOE Joint		
	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		
COMMENT	On Apr 20, 2001 this sequence version replaced gi:7711893.		
	-----Genome Center		
	Center: Joint Genome Institute		
	Center Code: JGI		
	Web site: http://www.jgi.doe.gov		

```

Project Information
Center Project Name: 671130, H460
Center clone name: CITE-H1_2113H21
-----
Summary Statistics
Consensus quality: 140723 bases at least Q40
Consensus quality: 148960 bases at least Q30
Consensus quality: 151290 bases at least Q20
Estimated insert size: 98000; pulse field gel estimation
Estimated insert size: 157572; sum-of-contigs estimation
Quality coverage: 7.39 in Q20 bases; pulse field gel estimation
Quality coverage: 4.59 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently

```


REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 181566)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
Unpublished
2 (bases 1 to 181566)
DOE Joint Genome Institute.
Direct Submission
Submitted (08-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 25, 2001 this sequence version replaced gi:11178071.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 594567
Center clone name: RPCI-11_490B18

Summary Statistics
Consensus quality: 153772 bases at least Q40
Consensus quality: 165049 bases at least Q30
Consensus quality: 169147 bases at least Q20
Estimated insert size: 182760; agarose-fp estimation
Estimated insert size: 178166; sum-of-contigs estimation
Quality coverage: 6.17 in Q20 bases; agarose-fp estimation
Quality coverage: 6.33 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 35 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1410: contig of 1410 bp in length
1411 1510: gap of unknown length
1511 2567: contig of 1057 bp in length
2568 2667: gap of unknown length
2668 4770: contig of 2103 bp in length
4771 4870: gap of unknown length
4871 6032: contig of 1162 bp in length
6033 6132: gap of unknown length
6133 7298: contig of 1166 bp in length
7299 7398: gap of unknown length
7399 8846: contig of 1448 bp in length
8847 8946: gap of unknown length
8947 10891: contig of 1945 bp in length
10892 10991: gap of unknown length
10992 12950: contig of 1959 bp in length
12951 13051: gap of unknown length
13051 14575: contig of 1525 bp in length
14576 14675: gap of unknown length
14676 16310: contig of 1635 bp in length
16311 16410: gap of unknown length
16411 19161: contig of 2751 bp in length
19162 19261: gap of unknown length
19262 21508: contig of 2247 bp in length
21509 21608: gap of unknown length
21609 24650: contig of 3042 bp in length
24651 24750: gap of unknown length
24751 26355: contig of 1605 bp in length
26356 26455: gap of unknown length
26456 28151: contig of 1696 bp in length
28152 28251: gap of unknown length
28252 30702: contig of 2451 bp in length
30703 30802: gap of unknown length
30803 34404: contig of 3602 bp in length
34405 34504: gap of unknown length
34505 37720: contig of 3216 bp in length
37721 37820: gap of unknown length
37821 40186: contig of 2366 bp in length
40187 40286: gap of unknown length

* 40287 42960: contig of 2674 bp in length
* 42961 43060: gap of unknown length
* 43061 46345: contig of 3285 bp in length
* 46346 46445: gap of unknown length
* 46446 50669: contig of 4224 bp in length
* 50670 50769: gap of unknown length
* 50770 53998: contig of 3229 bp in length
* 53999 54098: gap of unknown length
* 54099 58292: contig of 4194 bp in length
* 58293 58392: gap of unknown length
* 58393 64153: contig of 5761 bp in length
* 64154 64253: gap of unknown length
* 64254 71527: contig of 7274 bp in length
* 71528 71627: gap of unknown length
* 71628 78709: contig of 7082 bp in length
* 78710 78809: gap of unknown length
* 78810 87047: contig of 8238 bp in length
* 87048 87147: gap of unknown length
* 87148 94964: contig of 7817 bp in length
* 94965 95064: gap of unknown length
* 95065 102106: contig of 7042 bp in length
* 102107 102206: gap of unknown length
* 102207 108793: contig of 6587 bp in length
* 108794 108893: gap of unknown length
* 108894 117724: contig of 8831 bp in length
* 117725 117824: gap of unknown length
* 117825 132019: contig of 14195 bp in length
* 132020 132119: gap of unknown length
* 132120 153517: contig of 21398 bp in length
* 153518 153617: gap of unknown length
* 153618 181566: contig of 27949 bp in length.

FEATURES
Location/Qualifiers
source
1. 181566
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-490B18"
/clone_lib="RPCI human BAC library 11"
BASE COUNT 47714 a 40793 c 39929 g 49716 t 3414 others
ORIGIN

Query Match 100.0%; Score 1173; DB 2; Length 181566;
Best Local Similarity 100.0%; Pred. No. 5e-162;
Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgtggctgccacggttctccagcaagacagtgacagtgcctccttgccacagaccacc 60
|||||
Db 140275 ATGTGGCTGCCACGGTCTCCAGCAAGACAGTGACAGTGTCTCTCTGGCACAGACGACC 140216
QY 61 tgcctcctgctcttctcatcatctcccgccagggccctcatcccgagccggcgaggat 120
|||||
Db 140215 TGCCTCCTGCTCTTCATCATCTCCCGGCCAGGGCCCTCATCCCGCGGGCGAGGAT 140156
QY 121 cgtgtgcacgtgctggtgctgtctcctgtggcgctcgggctcatccttcttgggagcagctc 180
|||||
Db 140155 CGTGTGCACGTGCTGCTGTCTCTCGTGGCGCTCGGGCTCATCTCTTGTGGCGAGCTC 140096
QY 181 ttcagccagcaccgccgagcagctcttctacatgatggagccggcgagcctgatgccttacc 240
|||||
Db 140095 TTCAGCCAGCACCCCGGCGCAACGCTGCATGGAGCCGCGGTGGCATGTGTGGACGACC 140036
QY 241 ctgtgcagggcgagcgccgcaacgctgcacatggccgtggcgagcctgatgccttacc 300
|||||
Db 140035 CTGTGCGAGGGCGAGCGCGGCAACGCTGCATGGCGCTGCGGACCTGATCGCTCTATC 139976
QY 301 tttttgtgcgacatggacgtgtttgatgcctacatgccacagagccgaacacctgtccgcc 360
|||||
Db 139975 TTTTGTGCGACATGGACGTGTTTGTATGCTATACATGCCACAGAGCCGAACCTGTCCGCC 139916
QY 361 tttttcaactgggcaacgagccgagcggcgctgtgtcgcgcgcgcgcctgcagcgcccttccc 420
|||||
Db 139915 TTTTCAACTGGGCAACGAGCGCGGCTGTGTCTCGCCGCCCGCTGTCAGCGCCTTCCC 139856

QY 421 cgaggcaccatcagcaagcaggacgtatgcaagacactgtgcacgcgccagccattcagg 480
|||||
Db 139855 CGAGGCACCATCAGCAAGCAGGACGTAAGCAACACTGTGCACGCGGCAGCCAITCAGC 139796
QY 481 ctggcccccggagcctgcgcgtcctctacagcaacgtggtgctcaaggaggtgcgtctcttc 540
|||||
Db 139795 CTGGCCCCGGAGGCGCTGCCGCTCCTACAGCCACGTTGGTCTCAAGGAGGTGGCTTCTTC 139736
QY 541 aacctgcagtgctctacccgctgctcagcgaccccgcgctcaaacctgcgcatcgtgcac 600
|||||
Db 139735 AACCTGCAGGTGCTCTACCCGCTGCTCAGCGACCCCGCGCTCAACCTGGCATCGTGCA 139676
QY 601 ctggtgcgcgaccccgccgctgctgcgtcctcccgaggcggggcccgcatactggca 660
|||||
Db 139675 CTGGTGGCGGACCCCGCGGCGCTGCTGCGCTCCCGGAGGCGGCGGCGCGGATCGTGGCA 139616
QY 661 cgcgacaacggcatcgtgctgggacccaacggcaagtgggtgagggcgacccctcacctg 720
|||||
Db 139615 CGCGACAACGGCATCGTGTGGGACCAACGGAAGTGGTGGAGGCGGACCCCTCACCTG 139556
QY 721 cgctgattcgcgaggtgtgcgcagccacgtgcgcacgtgcgcacgtgcgcacgtgcgcac 780
|||||
Db 139555 CGCCTGATTTCGGAGGTGTGCGCGAGCCACGTCGCGCATCGCGAGGCGGCGGCGGCGGCGG 139496
QY 781 cgcgcaccccttcctgcgcggccgctacccgctggtgcgttcgaggaacctggcgggag 840
|||||
Db 139495 CGGCCACCCCTTCCTGCGGCGCGCTACCGCCTGTTGCGCTTCGAGGACCTGGCGGCGGCGG 139436
QY 841 ccgctggcagagatccgcgcactctacgccttcacggcctgcacccctcacgcccagctc 900
|||||
Db 139435 CCGCTGGCAGAGATCCGCGCACTCTACGCCCTTCACCGGCGCTGACCCCTCACGCCACAGCTC 139376
QY 901 gaggcctggatccacaacatcacccacgggtcggggatcggaagcccaatcgagggccttc 960
|||||
Db 139375 GAGGCCTGGATCCACAACATCACCCACGGGTGCGGATCGGAAGCCCAATCGAGGCCITC 139316
QY 961 catacttcgttaggaatgcgcgaacgtctccagggcctggcgccacgcgttgcccttc 1020
|||||
Db 139315 CATACTTCGTAGGAATGCGCGCAACGTCCTCCAGGCCCTGGCGGCCACGCGTIGCCCTTC 139256
QY 1021 actaagatcctgcgctgcaggaggtgtgcccggcgcgctgcagctgctgggctaccgg 1080
|||||
Db 139255 ACTAAGATCCTGCGGTGCAGGAGGTGTGCGCGGCGCGCTGCAGCTGCTGGGCTACCGG 139196
QY 1081 cctgttactctcggaccagcagcgtgacctcacccctgacctggtgctgccacgaggc 1140
|||||
Db 139195 CCTGTGFACTCTGCGGACCAAGCGGTGACCTCACCCCTGGATCTGCTGCTGCCACGAGGC 139136
QY 1141 ccagaccacttcagctgggcacatgcctgactga 1173
|||||
Db 139135 CCAGACCATTTCAGCTGGGCATCGCCTGACTGA 139103

RESULT 7
AC009163
LOCUS
DEFINITION Homo sapiens chromosome 16 clone RP11-77K12, WORKING DRAFT
AC009163 HTG 04-MAY-2000
SEQUENCE, 51 unordered pieces.
AC009163
VERSION AC009163.4 GI:7689976
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 179065)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 179065)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission

JOURNAL
COMMENT

Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On May 4, 2000 this sequence version replaced gi:7239429.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 436185
Center clone name: RPCI-11_77K12

Summary Statistics
Consensus quality: 132269 bases at least Q40
Consensus quality: 157541 bases at least Q30
Consensus quality: 162978 bases at least Q20
Estimated insert size: 63060; agarose-fp estimation
Estimated insert size: 174065; sum-of-contigs estimation
Quality coverage: 3.87 in Q20 bases; agarose-fp estimation
Quality coverage: 1.4 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 51 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1020: contig of 1020 bp in length
1021 1021: gap of unknown length
1121 2488: contig of 1368 bp in length
2489 2588: gap of unknown length
2589 4067: contig of 1479 bp in length
4068 4167: gap of unknown length
4168 5211: contig of 1044 bp in length
5212 5311: gap of unknown length
5312 6416: contig of 1105 bp in length
6417 6517: gap of unknown length
6517 7887: contig of 1371 bp in length
7888 7987: gap of unknown length
7988 9088: contig of 1101 bp in length
9089 9189: gap of unknown length
9189 10211: contig of 1023 bp in length
10212 10311: gap of unknown length
10312 11713: contig of 1402 bp in length
11714 11813: gap of unknown length
11814 13073: contig of 1260 bp in length
13074 13173: gap of unknown length
13174 14854: contig of 1681 bp in length
14855 14954: gap of unknown length
14955 16909: contig of 1955 bp in length
16910 17010: gap of unknown length
17010 18668: contig of 1659 bp in length
18669 18768: gap of unknown length
18769 20273: contig of 1505 bp in length
20274 20373: gap of unknown length
20374 21426: contig of 1053 bp in length
21427 21526: gap of unknown length
21527 22910: contig of 1384 bp in length
22911 23010: gap of unknown length
23011 25171: contig of 2161 bp in length
25172 25271: gap of unknown length
25272 27706: contig of 2435 bp in length
27707 27806: gap of unknown length
27807 29083: contig of 1277 bp in length
29084 29183: gap of unknown length
29184 31233: contig of 2050 bp in length
31234 31333: gap of unknown length
31334 33854: contig of 2521 bp in length
33855 33954: gap of unknown length
33955 36107: contig of 2153 bp in length
36108 36207: gap of unknown length
36208 38156: contig of 1949 bp in length

QY 1141 ccagaccacttcagctggcgatcgccctgactga 1173
 |||
 Db 167876 CCAGACCAC^{TTC}AGCTGGGCATCGCCTGACTGA 167908

RESULT	8
AF219990	
LOCUS	AF219990 2544 bp mRNA PRI 26-OCT-2000
DEFINITION	Homo sapiens corneal N-acetylglucosamine-6-O-sulfotransferase (CHST6) mRNA, complete cds.
ACCESSION	AF219990
VERSION	AF219990.1 GI:11023145
KEYWORDS	
SOURCE	human.
ORGANISM	Homo sapiens ⁴ Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 2544)
AUTHORS	Akama,T.O., Nishida,K., Nakayama,J., Watanabe,H., Fujiwara,I., Nakamura,T., Dota,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S., Ozaki,K., Kinoshita,S., Shimomura,Y., Tanigami,A. and Fukuda,M.N.
TITLE	Macular corneal dystrophy type I and type II are caused by distinct mutations in a new sulphotransferase gene
JOURNAL	Nat. Genet. 26 (2), 237-241(2000)

BASE COUNT	460 a	799 c	733 g	552 t
ORIGIN				
Query Match		78.8%	Score 924.6;	Length 2544;
Best Local Similarity		87.8%	Pred. No. 3.3e-125;	
Matches 1021;	Conservative	0;	Mismatches 139;	Indels 3;
				Gaps 1;

Qy 1 atgtggctgccagggttctccagcagacacagtgaacagtgcctctctctggtgcacagaccacc 60
|||||
Dd 693 AIGTGGCTGCCGCGGTCTCCAGCACACAGCAGTGACC GCGCTCTCTGGCGCAG -- ACC 749
|||||
Qy 61 tgcctcctgctcttcattcatctcccggccaggggccccctcatcccccaagccggcgggaggat 120
|||||

Db	750	TTCCNCTCCTCTCTTCTGTTTCCCGGCCAGGGCCCTCGTCCCCAGCAGGGCGGAGGCG	809
QY	121	cgtgtgcacgtgctggtgtgctcctcgtggcgctcgggctcatctcttggccagctc	180
Db	810	CGCGTGCATGTGCTGGTGTCTCCTCGTGGCGCTCGGGCTCGTCTTCGTGGGCCAACTC	869
QY	181	ttcagcagcaccocgacgtcttctacctgatggagcccgctggcatgtgtggaccacc	240
Db	870	TTCAACCAGCACCCCGACGCTTCTACCTAATGGAGCCCGGCTGGCAGCTGTGGACCAAC	929
QY	241	ctgtgcagggcagcgcggcaacgctgcacatggcctgcccgtgcggaacctgatgcgtctatc	300
Db	930	CTGTCCGAGGGCAGCGCCGCAACGCTGCACATGGCTGTGCGCAGCTTGGTGCCTCCGTC	989
QY	301	ttttgtgcacatggacgtgtttgatgcctacatgccacagagccgaaacctgtccgc	360
Db	990	TTCTGTGTGCACATGGACGTGTTTGTATGCTATCTGCCTTGGCGCCGCAACCTGTCCGAC	1049
QY	361	ttttcaactgggcaacgagccgcgcgctgtgtctgcgcgcgcctgcagcgccttccc	420
Db	1050	CTCTTCCAGTGGCCGTGAGCCGCTGCACGTGTCTGTGCTATCTGCCTTGGCGCCGCAACCTGTCC	1109
QY	421	cgaggcaacctcagcaagcaggacgtatgcaagacacactgtgcagcggcgagccattcagc	480
Db	1110	CGAGGGGCATCAGCAGCGAGGCCGTGTGCAAGCCACTGTGCGCGGGCAGTCCCTCACC	1169
QY	481	ctggccggggagccctgcgcgtcctctacagccacactgtgtcgaaggaggtgcgtctcttc	540
Db	1170	CTGGCCCGGGAGGCGCTGCCGCTCCTACAGCCACGTGTGCTCAAGGAGGTGCGCTTCTTC	1229
QY	541	aacctgcaggtgctctacccgctgctcagcagcccccgcctcaacctgcacatcgtgcac	600
Db	1230	AACTTGCAGGTGCTCTACCGGCTGCTCAGCGACCCCGCGCTCAACTACGCATCGTGCAC	1289
QY	601	ctggtgcgcgaccccgccggtgctgcgtcccgaggagcgccgggcccgcatactggca	660
Db	1290	CTGGTGGCGACCCCGCGGCGCTGTGCGCTCCCGGGAGCAGACAGCCAAAGGCTCTGGCG	1349
QY	661	cgcgacaacggcatcgtctgtgggcaccaaocggcaagtgggtggagccgacccctcacctg	720
Db	1350	CGTGACAACGGCATCGTCTGGGCACCAACCGGCACGTGGGTGGAGGCGGACCCCGGCGTG	1409
QY	721	cgctgattcgcgaggtgtgcgcgacccactgtgcgcctcgcgcatcgcgagccgcccacactcaag	780
Db	1410	CGCGTGGTGGCGAGGTGTGCGGTAGCCACGTACGCATCGCCGAGGCGCCGCCACACTCAAG	1469
QY	781	ccgcacaccttctcgcgcgcgctacccgctggtgcgttcgaggaacctggcgcgaggag	840
Db	1470	CCGCCACCGTTCTGCGCGGCGCTACCGCGCTGTTGCGCTTCGAGGACCTGGCGCGGGAG	1529
QY	841	ccgctggcagagatccgcgcactctacgccttaacccggcctgacctcagccacagctc	900
Db	1530	CCGCTGGCAGAAATCCGTGCGCTCTACGCTTCACCTGGGGCTCAGTCTACGCCACAGCTC	1589
QY	901	gaggcctggaatccacaacatcaccacgggtcggggatcggcaagccaaatcgaggccttc	960
Db	1590	GAGGCCCTGGATCCATAACATACACCCACGGAICTGTGGACCTGGTGGCGGCCCGGAAGCGCTC	1649
QY	961	catacttgcttaggaatgcgcgcaactctacgccttaacccggcctgacctcagccacagctc	1020
Db	1650	AAGACTTCGTCCAGGAATGCGCTCAACGCTCTCCAGGGCTTGGCGCATGCGCTGCCCTTT	1709
QY	1021	actaagatacctgcgcgtgcaggaggtgtgcgcgcgcgcgcgtgcagctgctgggctaccgg	1080
Db	1710	GCCAAAGATCCGCGCTGCAGGAACCTGTGGCGTGTGCGCTGCAGTGTGGGCTACCGG	1769
QY	1081	cctgtgtactctgcggaccagcagcgtgacctaacctcctggatctggtgtgccacagggc	1140
Db	1770	CCTGTGTACTCTGAGGACGAGCAGCGCAACCTCGCCCTTGTATCTGTGTGCTGCCACGAGGC	1829
QY	1141	ccagaccacttcagctgggcatc	1163
Db	1830	CTGAACGGCTTCACTTGGGCATC	1852

QY 481 ctggcccgaggcctgcgcgtctctacacccacgacgctggtgctcaaggagggtgcctcttc 540
Db 584 CTGGCTCAGGAAGCCTGCAGCTCTATAGTCACGTCTGCTCAAGGAGGTCGCTCTTT 643
QY 541 aacctgcaggtgctctacccgcgtcagcgaccccgctcaacctggcagc 600
Db 644 AACCTACAGGTGCTCTACCCGCTGCTCAGCGACCCCTCGCTCAACCTGCGCATCGTGAC 703
QY 601 ctggtgcgaccccgcggtgctgcctcagcgaccccgctcagcgagcgcggtgac 660
Db 704 CTAGTGGCGACCCCGCGGCGTGTGCTCCCGAGAGCAGACAGCCAAAGCGCTGGCA 763
QY 661 cgcgacaaacggcatcgtgctgggacacaaacggcaagtgggtggagggccgacctcacctg 720
Db 764 CGGACAAATGGCAATCGTCTGGGTACCAACGGCACGTGGGTGGAGCGGACCCCGGCTG 823
QY 721 cgcctgattgcgaggtgtgcgcagccacgacgtgcgcatcgccgagggccgacacacaa 780
Db 824 CGCGTGGTCAACGAGGTATGCCGAGCCCATGTGCGCATGCGAGAGCGAGCCTTGACACA 883
QY 781 cgcgcaacctctcgcgcggcgctacccgctggtgcgcttcgagggagcggcgag 840
Db 884 CGCGCGCCCTCTGTCAAGATCGGTACCGCTGGTGGCTACGAGGATCTGGCCGGAC 943
QY 841 cgcgtggcagagatccgcgcactctacgcttcacggcctgacccctcagccacagctc 900
Db 944 CCACTCACCGTAATCCGTGAACCTATGCTTACCGGCTGCTGCTCAGCGCGCAGCTC 1003
QY 901 gaggcctggatccacaaacataccacacgggtggtgggacggaagccaatcgaggccttc 960
Db 1004 CAGACTTGGATCCACAAATATACCGCATGTTTCAGGGCAGGCGCGCGGTGAAGCCTTC 1063
QY 961 catacctgtctaggaatgcgcgcaacgtctccagccctggcgccacggttgcccttc 1020
Db 1064 AAGACCAATCCAGGATGCGCTCAGTGTATCCAGGCGCTGGCGCCACAGCTGCCCCITT 1123
QY 1021 actaagatcctgcgcgtgcagaggtgtgcgcggcgctgcagctgcagctggtggctacc 1080
Db 1124 GCCAAGATCGCCGGTCCAGGAACGTGTGCGGGGTGCACTGCAGCTGCTGGGTACCGG 1183
QY 1081 cctgtgtactctgcggaccagcagcgtgacctacacctggatctggtgtgccacgagc 1140
Db 1184 TCTGTGCTATCGGAGCTTGAGCAAGGGACCTCTCTGTGGACCTCTGCTGCTGCTGCTGCTG 1243
QY 1141 ccagaccacttcagctgggcatcgcc 1166
Db 1244 ATGGACAGTTCAAGTGGGCATCGTC 1269

RESULT 14
AF131235
LOCUS AF131235 2032 bp mRNA PRI 10-AUG-1999
DEFINITION Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA,
complete cds.
ACCESSION AF131235
VERSION AF131235.1 GI:4927113
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2032)
AUTHORS Bistrup,A., Bhakta,S., Lee,J.K., Belov,Y.Y., Gunn,M.D., Zuo,F.R.,
Huang,C.C., Kannagi,R., Rosen,S.D. and Hemmerich,S.
Sulfotransferases of two specificities function in the
reconstitution of high endothelial cell ligands for L-selectin
J. Cell Biol. 145 (4), 899-910 (1999)
JOURNAL 99264336
MEDLINE 10330415
PUBMED 2 (bases 1 to 2032)
REFERENCE Bistrup,A., Tangemann,K., Bhakta,S., Lee,J.-K., Belov,Y.Y.,
AUTHORS Gunn,M.D., Zuo,F.-R., Huang,C.-C., Kannagi,R., Rosen,S.D. and
Hemmerich,S.

TITLE Direct Submission
JOURNAL Submitted (25-FEB-1999) Respiratory Diseases, Roche Bioscience,
3401 Hillview Avenue, Palo Alto, CA 94304, USA
FEATURES
source Location/Qualifiers
1..2032
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="high endothelial"
/tissue_type="tonsil"
182..1342
/function="biosynthesis of endothelial ligands for
L-selectin"
/note="HEC-GlcNAc-6-St"
/codon_start=1
/product="N-acetylglucosamine 6-O-sulfotransferase"
/protein_id="AAD33015.1"
/db_xref="GI:4927114"
/translation="MLLPKMKLLFLVSQMAILALFFHMYSHNLISSLSMAQPERMH
VLVLSWRSGSFVGLFGQHPDVFYLMPEPAWHVMTFKQSTAWMLHMAVRDLTRAVE
LQDMSVFDAYMEPGRRQSSLFQWNSRALCSAPACDIIPDEIIPRAHCRLLCSQOP
FEWVEKACRSYSHVVLKEVREFNLSLYPLKDPNLSLHVLHVRDPRAVRSRRTK
GDLMDSRIVMGQHEQKLKEDQPYVMQVICSQLEIYNTIQSLPKALQERILLVRY
EDLARAPVAOTSMYEFVGLFPLPLQTVHNTIRKGMGDHAFHTNARDALNVSQAW
RWSLPYEKVSRLQKACGDAMNLLGYRHRVSRSEQRNLLDLSTWTVPYEQIH"
BASE COUNT 468 a 569 c 490 g 505 t
ORIGIN

Query Match 34.1%; Score 400; DB 9; Length 2032;
Best Local Similarity 63.6%; Pred. No. 3.3e-49;
Matches 645; Conservative 0; Mismatches 360; Indels 9; Gaps 2;

QY 118 gatcgtgtgcagtgctggtgctgctctcctcctggtggcgtcgtcgtcctctcttggccag 177
Db 302 GAGCGCATGCACGTGCTGTTCTGTCTCTCTGCTGGCGCTGCTGCTCTCTCTTTGTGGGCGAG 361
QY 178 ctcttcagccacaccccgacgtctctctacctgatggagcccgctggtgcatgtgtggacc 237
Db 362 CTCTTTGGCGACACCCAGATGTTTCTACCTGATGGAGCCCGCTGGCATCTGCGGATCTGATG 421
QY 238 accctgtgcagggcagcgcgcgcaacgctgcacatggcgtgctgcgacacctgatgcgtct 297
Db 422 ACCTTCAAGCAGACGACCCGCTGGATGCTGCACATGGCTGCTGGGATCTGATACGGGCC 481
QY 298 atctttttgtgcagacatggacgtgtgtttgatgcctacatgccacagag---cggaaacctg 354
Db 482 GTCTTCTTGTGCGACATGAGCGTCTTTGATGCTTACATGSAACCTGGTCCCGGAGACAG 541
QY 355 tcggcctttttcaactgggcaacgagcgcgcgtgtgctgcgcgcgcgcgcgcgcgcgcgcgcgc 414
Db 542 TCCAGCCTCTTTCAGTGGGAGAACAGCGCGGCCCTGTGTCTGCACTGCTGCTGCTGCTGCTG 601
QY 415 ttccccgagggcaccatcagcaagcagcagcgtatgcaagacacactgtgcacgcgagcagcca 474
Db 602 ATCCACACAAGATGAATCATCCCGCGGCTCACTGCAGGCTCCTGTGCACTCAACACAGCCC 661
QY 475 ttccagcctggcccgaggaggcctgcgcctcctacagccagcagcagcagcagcagcagcagc 534
Db 662 TTTGAGGTGTGGAGAGGCGCTGCTTACAGCCACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 721
QY 535 ttcttcaacctgcaggtgctctacacgctgctcagcagcccgcgctcaacctgcgcagc 594
Db 722 TTCTTCAACCTGCAGTCCCTCTACCGCTGCTGAAAGACCCCTCCCTCAACCTGCATATC 781
QY 595 gtgcacctgtgcgcgaccccgcggtgctgctgcctcccgaggagcgggggcccgata 654
Db 782 GIGCACCTGTCCGGGACCCCGCGGCGTGTTCCTGTTCCCGAGAACGACGACAAAGGAGAT 841
QY 655 ctggcagcgcaaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 714
Db 842 CTCATGATTGACAGTCGCAATTGTGATGGGCGCAGCATGACGAGAAACTCAAGAGGAGGAC 901
QY 715 cacctgcgcctgattcgcgcgaggtgtgtgccgcagcagcagcagcagcagcagcagcagc 774

Direct Submission
Submitted (25-FEB-1999) Respiratory Diseases, Roche Bioscience,
3401 Hillview Avenue, Palo Alto, CA 94304, USA
Location/Qualifiers
1..2032
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="high endothelial"
/tissue_type="tonsil"
182..1342
/function="biosynthesis of endothelial ligands for
L-selectin"
/note="HEC-GlcNAc-6-St"
/codon_start=1
/product="N-acetylglucosamine 6-O-sulfotransferase"
/protein_id="AAD33015.1"
/db_xref="GI:4927114"
/translation="MLLPKMKLLFLVSQMAILALFFHMYSHNLISSLSMAQPERMH
VLVLSWRSGSFVGLFGQHPDVFYLMPEPAWHVMTFKQSTAWMLHMAVRDLTRAVE
LQDMSVFDAYMEPGRRQSSLFQWNSRALCSAPACDIIPDEIIPRAHCRLLCSQOP
FEWVEKACRSYSHVVLKEVREFNLSLYPLKDPNLSLHVLHVRDPRAVRSRRTK
GDLMDSRIVMGQHEQKLKEDQPYVMQVICSQLEIYNTIQSLPKALQERILLVRY
EDLARAPVAOTSMYEFVGLFPLPLQTVHNTIRKGMGDHAFHTNARDALNVSQAW
RWSLPYEKVSRLQKACGDAMNLLGYRHRVSRSEQRNLLDLSTWTVPYEQIH"

Fri Feb '1 20:20:55 2002

us-09-593-828-4.rge

Page 19

Search completed: January 31, 2002, 18:41:36
Job time: 12745 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:45:39 ; Search time 21.68 Seconds
(without alignments)
404.810 Million cell updates/sec

Title: US-09-593-828-8
Perfect score: 2051
Sequence: 1 MWLPRESSKTVIVLLAQTT.....LTLDLVLPGRGPDHFSWASPD 390

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1008	49.1	386	4	US-09-045-284A-2	Sequence 2, Appli
2	665	32.4	483	3	US-09-263-023-2	Sequence 2, Appli
3	656.5	32.0	484	3	US-09-263-023-4	Sequence 4, Appli
4	577.5	28.2	479	2	US-08-899-514-2	Sequence 2, Appli
5	540.5	26.4	458	2	US-08-655-878-2	Sequence 2, Appli
6	105	5.1	359	3	US-09-150-133-11	Sequence 11, Appli
7	105	5.1	359	3	US-09-150-141-11	Sequence 11, Appli
8	105	5.1	359	4	US-09-374-493-11	Sequence 11, Appli
9	105	5.1	359	4	US-09-374-824-11	Sequence 11, Appli
10	105	5.1	359	4	US-09-374-492-11	Sequence 11, Appli
11	90.5	4.4	380	3	US-09-150-133-9	Sequence 9, Appli
12	90.5	4.4	380	3	US-09-150-141-9	Sequence 9, Appli
13	90.5	4.4	380	4	US-09-374-493-9	Sequence 9, Appli
14	90.5	4.4	380	4	US-09-374-824-9	Sequence 9, Appli
15	90.5	4.4	380	4	US-09-374-492-9	Sequence 9, Appli
16	90.5	4.4	984	2	US-08-673-789-9	Sequence 9, Appli
17	86.5	4.2	376	3	US-09-150-133-7	Sequence 7, Appli
18	86.5	4.2	376	3	US-09-150-141-7	Sequence 7, Appli
19	86.5	4.2	376	4	US-09-374-493-7	Sequence 7, Appli
20	86.5	4.2	376	4	US-09-374-824-7	Sequence 7, Appli
21	86.5	4.2	376	4	US-09-374-492-7	Sequence 7, Appli
22	86.5	4.2	377	3	US-09-150-133-5	Sequence 5, Appli
23	86.5	4.2	377	3	US-09-150-141-5	Sequence 5, Appli
24	86.5	4.2	377	4	US-09-374-493-5	Sequence 5, Appli
25	86.5	4.2	377	4	US-09-374-824-5	Sequence 5, Appli
26	86.5	4.2	377	4	US-09-374-492-5	Sequence 5, Appli
27	86.5	4.2	566	2	US-08-484-993B-41	Sequence 41, Appli

28	86.5	4.2	566	2	US-08-484-158B-41	Sequence 41, Appli
29	86.5	4.2	566	2	US-08-484-596A-41	Sequence 41, Appli
30	86.5	4.2	566	2	US-08-480-150A-41	Sequence 41, Appli
31	86.5	4.2	566	3	US-08-458-731-41	Sequence 41, Appli
32	86.5	4.2	566	3	US-08-149-223A-41	Sequence 41, Appli
33	86.5	4.2	1711	3	US-08-369-822C-10	Sequence 10, Appli
34	86.5	4.2	1711	3	US-08-582-776C-10	Sequence 10, Appli
35	86.5	4.2	1711	3	US-08-434-831B-10	Sequence 10, Appli
36	83.5	4.1	390	4	US-09-108-020-36	Sequence 36, Appli
37	83	4.0	15281	2	US-08-471-119A-2	Sequence 2, Appli
38	81.5	4.0	829	4	US-08-444-818-69	Sequence 69, Appli
39	81.5	4.0	1786	4	US-08-444-818-54	Sequence 54, Appli
40	81.5	4.0	2261	4	US-08-444-818-66	Sequence 66, Appli
41	81.5	4.0	2436	4	US-08-444-818-75	Sequence 75, Appli
42	81.5	4.0	2772	4	US-08-444-818-89	Sequence 89, Appli
43	81.5	4.0	2894	2	US-08-466-975A-23	Sequence 23, Appli
44	81.5	4.0	2894	2	US-08-391-671A-23	Sequence 23, Appli
45	81.5	4.0	2894	3	US-08-467-902A-23	Sequence 23, Appli

ALIGNMENTS

RESULT 1
US-09-045-284A-2
; Sequence 2, Application US/09045284A
; Patent No. 6265192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107US1
; CURRENT APPLICATION NUMBER: US/09/045,284A
; CURRENT FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-045-284A-2

Query Match		49.1%;	Score 1008;	DB 4;	Length 386;
Best Local Similarity		52.3%;	Pred. No. 1.1e-102;		
Matches		206;	Conservative	56;	Mismatches 104; Indels 28; Gaps 7;
QY	14	LLAQTTCLLFIISRP-----	GPSSPAGGEDRVHVLVLSWSRSGSFLGQ	59	
Db	1	MLLPKMKLLFLVSQMAILALFFHMYSHNLSLSMAKQPERMHVVLSSWSRSGSFLGQ	60		
QY	60	LFSQHPDVFYLMPEPAWHVWTTLSQGSAAITLHMAVRDLMRISFLCDMDVFDAYM-PQSRNL	118		
Db	61	LFGQHPDVFYLMPEPAWHVWMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPGPRRQ	120		
QY	119	SAFFNWATSRALCSPACSAFPRGTISKQDVCKTLCRTQPFSLAREACRSYSHVVLKEVR	178		
Db	121	SSLFQWNSRALCSAPACDIIPQDEITPRAHCRLLCSQQPFVVEKACRSYSHVVLKEVR	180		
QY	179	FFNLQVLYPLSDPALNLRIVHLVRDPRAVLRSREAAGPILARDNGIVLGTN-GKWVEAD	237		
Db	181	FFNLQVLYPLSDPALNLRIVHLVRDPRAVLRSREAAGPILARDNGIVLGTN-GKWVEAD	240		
QY	238	PHLRLIREVCRSHVRIAEEATLKP-PPFLRGYRLVRFDLAREPLAEIRALYAFGLTL	296		
Db	241	QPYVMQVICQSQLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRMYEFGLEF	298		
QY	297	TPQLEAWIHNITHSGGIGKPTAFHTSSRNARNVSQAWRHLPFTKILRVQEVCAQALQL	356		
Db	299	LPHLQTVVHNITRGKMGD--HAFHTNARDALNVSQAWRWSLPYEKVSRLQKACGDAMNL	356		
QY	357	LGYRPVYSADQQORDLILDLVLPGRGPDHFSWASPD	390		

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,514
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: DANIEL E ALTMAN
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER: TOYAM21.001AUS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 714 760 0404
TELEFAX: 714 760 9502
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 479
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-899-514-2

Query Match 28.2%; Score 577.5; DB 2; Length 479;
Best Local Similarity 35.7%; Pred. No. 4.6e-55;
Matches 132; Conservative 65; Mismatches 134; Indels 39; Gaps 8;

QY 32 PSSPAGGEDRVHVLSSWRSGSFLGQSFQHPDVFYLMPEAWHVWTTLS---QGSAA 87
DB 123 PPRPAVAGPRRRHVLMMATRTGSSFVGEFFNQGNIFYLEFPLWHIERTVTFPPGGANAA 182
QY 88 TLHMAVRDLMRSIFLCMDVFDAY---MPSRNLSAFFNWTATSRALCSPACSAFPRGTI 144
DB 183 GSALIVRDVLKQFLCDLYVLEHFIITPLPEDHLTQFMFRGSSRSICEDPVCTPEVKVF 242
QY 145 SKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRFNQLVLYPLISDPALNLRIVHLVRD 204
DB 243 EKYHCKNRRCGPLNVTIAAEACRRKHEMALKAVRIQLEFLQPLAEDPRDLRVIQLVRD 302
QY 205 PRAVLRSREAAGPILARDNGIVLG---TNGKWVE-----ADPHLRLLIREVCRSHVRIA 254
DB 303 PRAVLASRMVA-----FAGKYKTWKWLDDEGQDGLREVEVQRLRGNCES-IRLS 351
QY 255 EAATLKPPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPQLEAWIHNIH---G 310
DB 352 AELGLRQPAWLGRYMLVRYEDVARGPLQKAREMYPFAGIPLTPQVEDWIQNTQAAHDG 411
QY 311 SGIGKPIEAFHTSSRNARNVSQAWRHALPFTKILRVQEVCAQALQLLGYRPPVYSADQORD 370
DB 412 SGI-----YSTQKNSSEQEKEKWRFSMPFKLAQVVQAPCGPAMRFLGYKLARDAALTN 464
QY 371 LTDLVLPRG 380
DB 465 RVSLLIEERG 474

RESULT 5
US-08-655-878-2
Sequence 2, Application US/08655878
Patent No. 5827713
GENERAL INFORMATION:
APPLICANT: FUKUTA, MASAKAZU
APPLICANT: HABUCHI, OSAMI
TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE:
STREET:

CITY:
STATE:
COUNTRY:
ZIP:
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/655,878
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE:
TELEFAX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 458
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-655-878-2

Query Match 26.4%; Score 540.5; DB 2; Length 458;
Best Local Similarity 34.3%; Pred. No. 5.1e-51;
Matches 124; Conservative 66; Mismatches 141; Indels 31; Gaps 8;

QY 36 AGGEDRVHVLSSWRSGSFLGQSFQHPDVFYLMPEAWHV---WTTLSQGSAA TLHMAV 93
DB 107 AAPEPRRHVLLMATRTGSSFVGEFFNQGNIFYLEFPLWHIERTVTFPPGGANAVGSAL 166
QY 94 --RDLMRSIFLCMDVFDAYM---PQSRNLSAFFNWTATSRALCSPACSAFPRGTISKQD 148
DB 167 VYROVLQQLLCLDLYILESFISPAPPEEHLTAALFRRGSSHSLSCEEPVCTPSLKKVFEKYH 226
QY 149 VCKTLCTROPFSLAREACRSYSHVVLKEVRFNQLVLYPLISDPALNLRIVHLVRDPRAV 208
DB 227 CKNRRCGPLNITLAAEACRRKQHMALKTVIRIQLEFLQPLAEDPRDLRLIQLVRDPRAV 286
QY 209 LRSREAAGPILARDNGIVLGTNGKWVE-----ADPHLRLLIREVCRSHVRIA EAATL 259
DB 287 LVSRMVA-----FSGKYESWKWAAEAGEAPLOEDEVEVQRLRGNCES-IRLSAELGL 335
QY 260 KPPFPLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPQLEAWIHNIHNSGIGKPIEA 319
DB 336 RQPRWLGRYMLVRYEDVARAPLKKALEMYRFAGIHTPTQVEEWIRANTQAP---QDSNG 392
QY 320 FHTSSRNARNVSQAWRHALPFTKILRVQEVCAQALQLLGYRPPVYSADQORDLTLDLVLPR 379
DB 393 IYSTQKNSSEQEKEKWRFSIPEFKLAQVVODACEPAMRFLGYKLASSAQELNRSLSL-LEE 451
QY 380 GP 381
DB 452 GP 453

RESULT 6
US-09-150-133-11
Sequence 11, Application US/09150133B
Patent No. 6060295
GENERAL INFORMATION:
APPLICANT: The Board of Regents of the University of Oklahoma
TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF

Query Match 4.4%; Score 90.5; DB 3; Length 380;
Best Local Similarity 19.3%; Pred. No. 0.16;
Matches 66; Conservative 43; Mismatches 108; Indels 125; Gaps 17;

QY	51	RSGSFELGQFSQHPDV	-----FYLMEPAW-----HVWTTLSQ-----SAATLHMAY	93
		:: : :	: : : :	
Db	79	RSGTTLMRAMLDAHPEVRCGEETRVIPRIILNRSQWKSEKEWNRIQQAGVTGEVINNAI	138	
QY	94	RDLMRSIFLCDMDVFDAYMPEQSRNLSAFFNWATSRALCSPACSAFFRGFTISKQDVCKTL	153	
Db	139	SSFIMEIM-----	-----VGHGDRAPRL	156
QY	154	CTRQPFSLAREACRSYSHVYLKEVRFNFNLQVLYPLLSDPALNLRIVHLVRDRPRAVLRSE	213	
Db	157	CNKDPPTMK-----SAVYLKE-----LFP-----NAKYLLMIRDGRATVNS--	192	
QY	214	AAGPILARD--NGIVLG----	TNGKWVEADPHLLKLIREVCSRSHVRIAEATLKPPPFRL	266
Db	193	-----IISRKVTITGFDLNDFFQCMTKWNA--	IQIMVDQCES--VGEKNCLK-----	236
QY	267	GRYRLVRFEDLAREPLAEIRALYAFTGLTLTPQLEAWIHNITHGSGIGKPIEAFHTSSRN	326	
Db	237	-----VYYEQVLVHPEAQMRRITEFLDI---	PWDDKVLH---HEQLIGKDLSLNSVERSS	285
QY	327	ARNYSQA-----	WRHALPFTKILRVQEVCAQALQLGYRP	361
Db	286	DQVVKPVNLDALIKWVGITPEDVADMDSV-APMLRRLGYDP	326	

```

RESULT 12
US-09-150-141-9
; Sequence 9, Application US/09150141B
; Patent No. 60711732
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.495
; CURRENT APPLICATION NUMBER: US/09/150,141B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 9
; LENGTH: 380
; TYPE: FRT
; ORGANISM: Caenorhabditis elegans
US-09-150-141-9

```

Query Match 4.4%; Score 90.5; DB 3; Length 380;
Best Local Similarity 19.3%; Pred. No. 0.16;
Matches 66; Conservative 43; Mismatches 108; Indels 125; Gaps 17;

QY	51	RGSSFLGQLFSQHPDV-----FYLMPEAW---HVWTTLSQG--SAATLHMVAV	93
		::: : : : : : : : : :	
Db	79	RSGETLMRAMLDHPEVRGCEETRVIPRIILNLSQWKSEKEWNRLQOAGVTGEVINNAI	138
QY	94	RDLMRISFLCDMDVFDAYMPQSRNLSAFFNWATSRALCSPACSAFPRGTISKQDVCKTL	153
Db	139	SSPIMEIM----- : : : : : :	156
QY	154	CTRQPEFSLAREACRSYSHVVLKEVREFNLQVLPYLLSDPALNLRIVHLVDPDRAVLRSRE	213
Db	157	CNKDPFTMK-----SAVYLKE-----LFP-----NAKYLLMIPTDGRATVNS--	192
QY	214	AAGPILARD---NGIVLG-----TNGKWVEADPHLRLIREVCRSHVRIAEATLKPPFRLR	266
Db	193	---ILSRKVTITGFDLNDFRQCMTKWNAA---IQINVDQCES---VGEKNCLK-----	236
QY	267	GRYLRVFEDLAREPLAEIRALYAFTGLTLPQLEAWIHNIHGSIGICKPIEAFHISRN	326
Db	237	---VYFEQLVLHPEAQMRRITEFLDI---PWDDKVLH---HEQLIGKDISLSNVERSS	285

QY 327 ARNYSQA-----WRHALPFTKILRYQEVCAGALQLLGYP 361
:
Db 286 DQVVKPVNLDAIKWGTIPEDVVADMSV-APMLRRLGYDP 326

```

RESULT 13
US-09-374-493-9
; Sequence 9, Application US/09374493
; Patent No. 6204016
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.546
; CURRENT APPLICATION NUMBER: US/09/374,493
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 9
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-374-493-9

```

Query Match 4.4%; Score 90.5; DB 4; Length 380;
Best Local Similarity 19.3%; Pred. No. 0.16;
Matches 66; Conservative 43; Mismatches 108; Indels 125; Gaps 17;

[illegible]

RESULT 14
US-09-374-824-9
; Sequence 9, Application US/09374824
; Patent No. 6207414
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820,547
; CURRENT APPLICATION NUMBER: US/09/374,824
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: January 31, 2002, 18:20:51 ; Search time 4771.45 Seconds
(without alignments)
7067.083 Million cell updates/sec

Title: US-09-593-828-3
Perfect score: 2044
Sequence: 1 ggctcgaggtocactgtgct.....aaaaaaaaaaaaaaaaaaaa 2044

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl:*
- 1: gb_ba:*
 - 2: gb_htg:*
 - 3: gb_in:*
 - 4: gb_om:*
 - 5: gb_ov:*
 - 6: gb_pat:*
 - 7: gb_ph:*
 - 8: gb_pl:*
 - 9: gb_pr:*
 - 10: gb_ro:*
 - 11: gb_sts:*
 - 12: gb_sy:*
 - 13: gb_un:*
 - 14: gb_vi:*
 - 15: em_ba:*
 - 16: em_fun:*
 - 17: em_hum:*
 - 18: em_in:*
 - 19: em_om:*
 - 20: em_or:*
 - 21: em_ov:*
 - 22: em_pat:*
 - 23: em_ph:*
 - 24: em_pl:*
 - 25: em_ro:*
 - 26: em_sts:*
 - 27: em_sy:*
 - 28: em_un:*
 - 29: em_vi:*
 - 30: em_htgo_hum:*
 - 31: em_htgo_inv:*
 - 32: em_htgo_rod:*
 - 33: em_htg_hum:*
 - 34: em_htg_inv:*
 - 35: em_htg_rod:*
 - 36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	1973	96.5	2170	9	AF176838	AF176838 Homo sapi
2	1774.6	86.8	3278	9	AF246718	AF246718 Homo sapi
C 3	1753.6	85.8	181566	2	AC025287	AC025287 Homo sapi
4	1742.6	85.3	159072	2	AC026419	AC026419 Homo sapi
5	1719	84.1	71503	9	AF219991	AF219991 Homo sapi
6	1628.8	79.7	179065	2	AC009163	AC009163 Homo sapi
7	1248.6	61.1	1462	9	AF176839	AF176839 Homo sapi
8	1101	53.9	3786	9	AF280086	AF280086 Homo sapi
9	1052	51.5	2544	9	AF219990	AF219990 Homo sapi
C 10	1046.4	51.2	159072	2	AC026419	AC026419 Homo sapi
11	1046.4	51.2	208185	2	AC009105	AC009105 Homo sapi
12	751.2	36.8	1989	10	AF176840	AF176840 Mus muscu
13	750.4	36.7	1740	10	AF176841	AF176841 Mus muscu
C 14	568	27.8	215647	2	AC068591	AC068591 Homo sapi
C 15	556.8	27.2	174187	2	AP002792	AP002792 Homo sapi
C 16	542.6	26.5	137499	2	AC015931	AC015931 Homo sapi
C 17	524.8	25.7	165762	2	AP001582	AP001582 Homo sapi
C 18	461.2	22.6	179065	2	AC009163	AC009163 Homo sapi
19	400	19.6	2032	9	AF131235	AF131235 Homo sapi
20	398.4	19.5	1333	9	AF149783	AF149783 Homo sapi
21	398.4	19.5	1992	9	AF280088	AF280088 Homo sapi
22	398.4	19.5	2011	9	AK026635	AK026635 Homo sapi
23	398.4	19.5	183133	2	AC010547	AC010547 Homo sapi
24	358.8	17.6	2201	10	AF109155	AF109155 Mus muscu
25	354	17.3	1926	10	AF131236	AF131236 Mus muscu
26	235.2	11.5	2156	6	AR071396	AR071396 Sequence
27	235.2	11.5	2156	6	E14937	E14937 Human mRNA
28	235.2	11.5	2731	9	AB012192	AB012192 Homo sapi
29	235.2	11.5	6961	9	AB017915	AB017915 Homo sapi
C 30	235.2	11.5	182677	2	AC073370	AC073370 Homo sapi
31	235.2	11.5	196465	9	AC022392	AC022392 Homo sapi
C 32	231	11.3	176752	2	AL590620	AL590620 Homo sapi
C 33	228.8	11.2	172176	2	AL450304	AL450304 Homo sapi
C 34	226.2	11.1	103567	9	HS124C6	HS124C6 Homo sapi
C 35	224.4	11.0	54666	9	AC073487	AC073487 Homo sapi
C 36	223	10.9	154962	2	AL358153	AL358153 Homo sapi
C 37	223	10.9	166820	2	AL590126	AL590126 Homo sapi
38	222	10.9	140210	2	AC002993	AC002993 Homo sapi
C 39	221.6	10.8	195382	2	AL445254	AL445254 Homo sapi
C 40	220.8	10.8	151068	9	AL133405	AL133405 Human DNA
C 41	220.6	10.8	148463	2	AC012140	AC012140 Homo sapi
42	220.6	10.8	207408	2	AC092872	AC092872 Homo sapi
43	220.6	10.8	213227	2	AC021027	AC021027 Pan trogl
C 44	220.4	10.8	57573	2	AC087509	AC087509 Homo sapi
45	220.4	10.8	182012	2	AC025590	AC025590 Homo sapi

ALIGNMENTS

RESULT 1

AF176838

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AF176838 2170 bp mRNA 22-SEP-1999
Homo sapiens intestine N-acetylglucosamine 6-O-sulfotransferase
(I-GlcNAc-6-ST) mRNA, complete cds.

AF176838.1 GI:5917705

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2170)

Lee,J.K., Bhakta,S., Rosen,S.D. and Hemmerich,S.
Cloning and characterization of a mammalian
N-acetylglucosamine-6-sulfotransferase that is highly restricted to
intestinal tissue

Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)

99423499

2 (bases 1 to 2170)

AUTHORS Lee, J.-K., Bhakta, S., Rosen, S.D. and Hemmerlich, S.
 TITLE Direct Submission
 JOURNAL Submitted (10-AUG-1999) Respiratory Diseases, Roche Bioscience,
 3401 Hillview Avenue, Palo Alto, CA 94304, USA
 FEATURES
 source Location/Qualifiers
 1. .2170
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /map="16q23.1"
 /tissue_type="intestine"
 1. .2170
 /gene="I-GlcNAC-6-ST"
 344. .1516
 /gene="I-GlcNAC-6-ST"
 /note="possibly involved in biosynthesis of mucosal
 ligands for L-selectin"
 /codon_start=1
 /product="N-acetylglucosamine 6-O-sulfotransferase"
 /protein_id="AAD56000.1"
 /db_xref="GI:5917706"
 /translation="MWLPFRSSKTVIVLLAQTCTLLFIIRPGPSSPAGGEDRVHV
 LVLSWRSGSFLQLFSQHPDVFYLMPEAWHVWTTLSQSAATHMAVRDLMRSIFL
 CMDVFDAYMPQSRNLSAFFNWTSRALCSPPACSAFPRGTISKQDVCKTLCTROPFS
 LAREACRSYSHVVLKEVRFNQLVLYPLLSDPALNRIHVLRDPRAVLRSREAAGPI
 LARDNGIVLGTNGKWEADPHLRITREVCRSHVRIAEATLKPPPELGRYRLVRFED
 LAREPLAETRALYAFTGLTLTPQLEAWITHNTHGSGIGKPIEAFTTSRNARNVSQLW
 RHALPFTKILRVQEVCAQALQLLGYRPVYSADQQORDLTLDLVLPGRDPHESWASPD"
 449 a 682 c 612 g 427 t
 BASE COUNT
 ORIGIN

[illegible]

Db 1786 AGTTACTGCTAAATTAAATTAAATGTGTGCCAGCGCGGGTGCGGTGCATGCCCTGTAA 1845

QY 1720 tcccagcattttgagaggctgagcggttgatcacctgaggtcaggattcaaaaccagc 1779

Db 1846 TCCAGCATTTTGAGAGGCTGAGCGGGTGGATCACCTGAGGTCAGGATCAAAACCAGC 1905

QY 1780 ctggccaacatagtgaacccccctctctactaaaaatgcaaaaattagtcggcggtgtg 1839

Db 1906 CTGGCCACATAGTGAACCCCTCTCTACTAAAAATGCAAAATTAGTCGGCGGTGGTG 1965

QY 1840 gcacactctgttaatcccagctacttaggaggtgaggtgggaaaaatcaacttggactcca 1899

Db 1966 GCACACTCCTGTAATCCCAGCTACTTAGAGGCTGAGGTGGGAGATCACTTGGACTCCA 2025

QY 1900 aaggtgaggttcagtaagctgaatcatgccactgccacctagcttgggtggcaaaagc 1959

Db 2026 AAGGTGAGGTTGCAGTAGCTGAATCATGCCACTGCACCCTAGCTTGGTGGCAAGC 2085

QY 1960 aaaaacttatcaaaaaataataataatttgttcaaaaagctcgcgaaaaaataaaa 2019

Db 2086 AAAACTCTATCAAAAAATAATAATAAATTGTTCAAAAGTCTGCCGAAAAAATAA 2145

QY 2020 aaaaaaataaaaaaataaaaaa 2044

Db 2146 AAAAAAATAAAAAAATAAAAAA 2170

RESULT 2

AF246718 3278 bp mRNA PRI 31-OCT-2000

LOCUS Homo sapiens intestinal GlcNAc-6-sulfotransferase (CHST5) mRNA, complete cds, alternatively spliced.

ACCESSION AF246718

VERSION AF246718.1 GI:11055254

KEYWORDS .

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 3278)

AUTHORS Akama,T.O., Nishida,K., Nakayama,J., Watanabe,H., Ozaki,K., Nakamura,T., Dota,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S., Fujiwara,T., Thonar,E.J., Shimomura,Y., Kinoshita,S., Tanigami,A. and Fukuda,M.N.

TITLE Macular corneal dystrophy type I and type II are caused by distinct mutations in a new sulphotransferase gene

JOURNAL Nat. Genet. 26 (2), 237-241 (2000)

MEDLINE 20472330

PUBMED 11017086

REFERENCE 2 (bases 1 to 3278)

AUTHORS Akama,T.O., Nishida,K., Nakayama,J., Watanabe,H., Ozaki,K., Nakamura,T., Dota,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S., Fujiwara,T., Thonar,E.J.-M.A., Shimomura,Y., Kinoshita,S., Tanigami,A. and Fukuda,M.N.

TITLE Direct Submission

JOURNAL Submitted (17-MAR-2000) Glycobiology Program, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA

FEATURES

Location/Qualifiers

1..3278

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="16"

/map="16q22: between D16S512 and D16S515"

1..3278

/gene="CHST5"

1..1396

/gene="CHST5"

/note="alternatively spliced"

1397..2632

/gene="CHST5"

/note="I-GlcNAc6ST; alternatively spliced"

/codon_start=1

/product="intestinal GlcNAc-6-sulfotransferase"

gene

5'UTR

CDS

/protein_id="AAG28023.1"

/db_xref="GI:11055255"

/translation="MGMRARVPKVAHSTRPPAARMWLPRFSSKIVVLLAQITCLL

LFIIISRPGPSPPAGGEDRVHVLVSSWSRGSFGLQFQSOHPDVFYLMFPAHWVWTL

SQGSAAALHMAVRDLMSIFLDCMDVEDAYMPQSRNLSAFFNWATSRALCSPPPACSAF

PRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVFFNLQVLYPLSDPALNLR

IVHLVRDPRAVLRREAAGPILARDNGIVLGTNGKWEADPHRLRLEFVCRSHVRIAE

AATLKPFPFLRGYRLVRFEDILAREPLAEIRALYAFGTGLTLPQLEAWIHNIHSGSI

GKPIEAFHTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQLLGYRPVYSADQORDL

TLDLVLPRGPDHFSWASPD"

BASE COUNT 686 a 988 c 866 g 738 t

ORIGIN

Query Match 86.8%; Score 1774.6; DB 9; Length 3278;

Best Local Similarity 98.8%; Pred. No. 4e-284;

Matches 1819; Conservative 0; Mismatches 19; Indels 3; Gaps 3;

QY 197 cggtggccccagccgcccgcgcatgttgctgccacggttctccagcaagacagtgacagtg 256

Db 1439 CGAAGGCCCCAGCGCCGCATGTGGTGGCCACGGTTCTCCAGCAAGACAGTGACAGTG 1498

QY 257 ctctctctggcacagaccacctgcctcctgctcttcatcatctcccgccagggccctca 316

Db 1499 CTCCTCCTGGCACAGACCACCTGCCTCCTGCTCTTATCATATCTCCGGGCCAGGCCCTCA 1558

QY 317 tcccagccggcgcgaggatcgtgtgcacgtgctgtgctgtgctgtgctgtgctgtgctg 376

Db 1559 TCCCCAGCCGGCGGAGGATCGTGTGCACGTGCTGGTGTCTCTGTTGGTGGCGCTCGGGC 1618

QY 377 tcaatcttcttgggcccagctcttccagccagcaccggacgttcttacctgatggagccc 436

Db 1619 TCATCCCTTCTTGGGCCAGCTCTTTCAGCCAGCACCCCGACGCTCTTCTACCTGATGGAGCCC 1678

QY 437 gcgtggcatgtgtggaccacctgtgcagggcagggcagggcagggcagggcagggcaggg 496

Db 1679 GCGTGGCATGTGTGGACCAACCTGTTCGCAGGGCAGCGGGCAACGCTGCACATGGCCGTG 1738

QY 497 cgcgacctgatgcgctctatcttttgcgacatggacgtgtttgatgcctaatgccca 556

Db 1739 CGCGACCTGATGCGCTCTATCTTTTGTGCGACATGGACGTGTTGATGCCTACATGCCA 1798

QY 557 cagagccgaaaacctgtccgcctttttcaactgggcaagccgcgcgtgtgctgcgcg 616

Db 1799 CAGAGCCGAAACCTGTCCGCCCTTTTCAACTGGGCAACGAGCCGCGCTGTGCTCGCG 1858

QY 617 cccgcctgcagcgcccttccccgaggccacctcagcaagcagggacgtatgcaagacactg 676

Db 1859 CCCGCCTGCAGCGCCTTTCCCGAGGCACCATCAGCAAGCAGGACGATGCAAGACACTG 1918

QY 677 tgcacgggcagccattcagcctggccggggagggcctgcgcgtcctctacagccacgtgtg 736

Db 1919 TGCACGGCGCAGCCATTTCAGCCTGGCCGGGAGGGCCTGCCGCTCTCTACAGCCACGTGGTG 1978

QY 737 ctcaaggaggtgcgcttcttcaacctgcaggtgctctacccgctgctcagcgaccccgcg 796

Db 1979 CTCAAGGAGGTGCGCTTCTTCAACCTGCAGGTGCTCTACCCGCTGTCTACCGGACCCCGCG 2038

QY 797 ctcaacctgcgcacgtgcacctggtgcgcgaccccgccgctgtgcgctcccgagg 856

Db 2039 CTCAACCTGCGCATCGTGCACCTGGTGGCGGACCCCGGCGCTGTCGCGCTCCCGGAG 2098

QY 857 gcggcgggcccgatactggcacgcgacaaacggcatcgtgtgtggcaccacgcaagtgg 916

Db 2099 GCGGCGGGCCCGATACTGGCACGCGACAACGGCATCGTGTGGGCACCAACGCAAGTGG 2158

QY 917 gtggaggccgaacctcacctgcgcctgattcggagggtgtgtgccgcagccacgtgcgcac 976

Db 2159 GTGGAGGCCGACCTCACTGCGCCTGATTGCGGAGGTGTGCCGAGCCACGTCGCATC 2218

QY 977 gccgaggccgcacactcaagccgcccaccttctctgcgcgcccgtacgcctgtgtgcgc 1036

Db 2219 GCCGAGGCCGCCACACTCAAGCCGCCACCCCTTCTGCGCGCGCGCTACCCGCTGGTGCGC 2278

QY 497 cgcgacctgatgcgctctatctttttgtgagacatggacgtgttttgatgcctacatgcca 556
Db 61679 CGCGACCTGATGCGCTCTAICTTTTGTGGACATGGACGTGTTTGATGCTACATGCCA 61738
QY 557 cagagccgaacctgtccgcctttttcaactgggcaacgagccgcgcgcgtgtgctgcgcg 616
Db 61739 CAGAGCCGAACCTGTCCGCCCTTTTCAACTGGCAACGAGCCCGCGCTGTGCTGCCG 61798
QY 617 cccgcctgcagccctttcccccagggcaccatcagcaagcaggaagcgtatgcaagacactg 676
Db 61799 CCCGCTGCAGCGCCTTTCCCGAGGACCATCAGCAAGCAGGACGTATGCAAGACACTG 61858
QY 677 tgcacgcggcagccattcagcctgcccgggagccctgcgcgtcctcactacagccactggtg 736
Db 61859 TGCACGCGGCAGCCATTTCAGCCTGGCCCGGAGGCCCTGCCCTCTACAGCCACGTGGTG 61918
QY 737 ctcaaggagggtgcgctttcttcaacctgcaggtgctctacccgtgctcagcgacccccgcg 796
Db 61919 CTCAAGGAGGTGCGCTTCTCAACCTGCAGGTGCTCTACCCGCTGCTCAGCGACCCCGCG 61978
QY 797 ctcaacctgcgcacatcgtgcacctggtgcgcgacccggcgccgtgctgcgcctcccgggag 856
Db 61979 CTCAACCTGCGCATCGTGCACCTGCTGCGGACCCCGCGCGCGTGTCTGCGCTCCCGGGAG 62038
QY 857 gcggcgcccgatactggcacgcgacacggcctgctgctgggcacccacacgcaagtgg 916
Db 62039 GCGGCGGGCCGATACTGGCACGCGACAAACGGCATCGTGTGGGCACCAACGCGCAAGTGG 62098
QY 917 gtggaggccgacacctcacctgcgcctgattgcgaggtgtgccgcagccacactgcgcac 976
Db 62099 GTGGAGGCGGACCCCTCACCTGGCCCTGATTGCGGAGGTGTGCCGCAGCCACGTGCGCAIC 62158
QY 977 gccgagcgccacactcaagccgcacaccttctcctgcgcggcgctacgcctgtgtgcgc 1036
Db 62159 GCCGAGGCGGCACACTCAAGCCGCCACCCCTTCTGTGCGCGCGGCTACCGCTGGTGCGC 62218
QY 1037 ttgcgagacctggcgcgggagccgtggcagagatccgcgcactctacgccttcaccgcgc 1096
Db 62219 TTCGAGACCTTGGCGCGGGAGCCGCTGGCAGAGATCCGCGCACTCTACGCCCTTCAACGCGC 62278
QY 1097 ctgacctcacgccacagctcgagggcctggatccacaacatcaccccaagggtcggggatc 1156
Db 62279 CTGACCCCTCAGCCCAAGCTCGAGGCGCTGGATCCACAACATCACCCACGGGTGCGGGATC 62338
QY 1157 ggcaagccaatcgaggcccttcatactctgtctaggaatgcgcgcgcaactctccacggcc 1216
Db 62339 GGCAAGCCAATCGAGGCGCTTCCATACTTCTGTCTAGGAATGCGCGCAACGTCTCCAGGCC 62398
QY 1217 tggcgccacgcgttgcccttcactaagatcctgcgcgtgcaggaggtgtgcgcgcgcgcgcg 1276
Db 62399 TGGCGCCACGCGTGGCCCTTCACTAAGATCCTGCGGTGCAGGAGGTGTGCGCGCGCGCG 62458
QY 1277 ctgcagctgctgggtaccggccctgtgtactctgcggaccagcagcgtgacactcacctg 1336
Db 62459 CTGCAGCTGCTGGGCTACCGGCCCTGTGTACTCTGCGGACCAAGCAGCGTGACCTCACCCCTG 62518
QY 1337 gatctggtgctgccacgagggccagaccacttcagctgggcatcgccctgactgagaactc 1396
Db 62519 GATCTGTTGCTGCCACGAGGCCAGACCCTTACGCTGGGCATCGCCTGACTGAGAACTC 62578
QY 1397 tgggacctagagcaagccccgaactgtgtgcaggccccagggaagcgactgcagtgtg 1456
Db 62579 TGGGCCCTTAGAGCAGGCCCGCAACTGTGTGCGCCAGGCCCTTCAACCGGGAGTTTGGGGTCC 62638
QY 1457 aaaaggagctggggcgcatggggaacaggtccctactatcaacccgggagtttggggtcc 1515
Db 62639 AGAGGAGCTGGGGCGCATGGGGAACAGAGTCCCTACTATCAACCGGGAGTTTGGGGTCC 62698
QY 1516 t-ccccctgaagtaagcaaggactgcagctttctttctctctctctctctctctctctct 1574
Db 62699 TCCCCCTGAAGTAGGCAAGGACTGCACGTTTCTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 62758

QY 1575 gagtcttctggagctgccttctcatcaggtgcactcttctcatgaaaaagcaactcttgccc 1634
Db 62759 GAGTCTTCTGGAGCTGCCTTCTCATCAGGTGCACCTCTCATGG-AAAGCAACCTCTTGCC 62817
QY 1635 claccttcttctggcgaggaagtaagttaactgctaaataataataatgtgtgcaggc 1694
Db 62818 CTACCTCTTCTGGCGCAGGGAGTAAGTTACTGCTAAATTAATTAATGTGCGCAGGC 62877
QY 1695 cgggtgcggtggtcctcatgcctgttaatccccagcattttgagaggtgagcggttgatca 1754
Db 62878 CGGGTGCCTGGTGCATGCCTGTAAATCCAGCATTTTGAAGGGCTGAGCGGGTGGATCA 62937
QY 1755 cctgaggtcagga-ttcaaaaccagcctggccaacatagtgaaaacccctctctactaaa 1813
Db 62938 CCTGAGGTACAGGAGTTCGAAACCCAGCCTGGCCAACATAGTGAACCCCTCTCTACTAAA 62997
QY 1814 aatgcaaaaattagtcgcggtgtgtggcacactcctgttaatcccagctacttaggagct 1873
Db 62998 AATGCAAAAATTAGTCCGGCGTGTGGTGCACACTCCTGTAAATCCAGCTACTTAGGAGGCT 63057
QY 1874 gagtgggaaaaatcacttggactccaaaagggtggaggttgaggttgagtaagctgaaatcatgcca 1933
Db 63058 GAGTGGGAGAAATCACTTGGACTCCAGAGGTGGAGGTGGAGTAAGCTGAGATCATGCCA 63117
QY 1934 ctgcacccctagcttgggtggcaaaagcaaaactctatcaaaaaataataataatttgt 1993
Db 63118 CTGCACCCCTAGCTTGGTGGTGGCAGAGCAAGACTCTATCAAAAAATAAGTAATAATTGT 63177
QY 1994 tcaaaagtccctgc 2006
Db 63178 TCAAAAGTCCTGC 63190

RESULT 5
AF219991
LOCUS AF219991 71503 bp DNA PRI 26-OCT-2000
DEFINITION Homo sapiens intestinal N-acetylglucosamine-6-O-sulfotransferase (CHST5) and corneal N-acetylglucosamine-6-O-sulfotransferase (CHST6) genes, complete cds.
ACCESSION AF219991
VERSION AF219991.1 GI:11023147
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 71503)
AUTHORS Akama,T.O., Nishida,K., Nakayama,J., Watanabe,H., Fujiwara,T., Nakamura,T., Dota,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S., Ozaki,K., Kinoshita,S., Shimomura,Y., Tanigami,A. and Fukuda,M.N.
TITLE Macular corneal dystrophy type I and type II are caused by distinct mutations in a new sulphotransferase gene
JOURNAL Nat. Genet. 26 (2), 237-241 (2000)
MEDLINE 20472330
PUBMED 11017086
REFERENCE 2 (bases 1 to 71503)
AUTHORS Akama,T.O. and Fukuda,M.N.
TITLE Direct Submission
JOURNAL Submitted (29-DEC-1999) Glycobiology Program, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA
FEATURES
Source Location/Qualifiers
1. 71503
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q22"
/clone="CITB-483K2"
358. 461
/rpt_family="L2"
/rpt_type=dispersed
repeat_region
repeat_region complement(462..771)
/rpt_family="Alu"
/rpt_type=dispersed

repeat_region	772. .945 /rpt_family="L2" /rpt_type=dispersed	repeat_region	/rpt_type=dispersed 15668. .15718 /rpt_family="Alu" /rpt_type=dispersed
repeat_region	1169. .1338 /rpt_family="MIR" /rpt_type=dispersed	repeat_region	15977. .16098 /rpt_family="MIR" /rpt_type=dispersed
repeat_region	complement(1553. .1804) /rpt_family="Alu" /rpt_type=dispersed	repeat_region	complement(17298. .17433) /rpt_family="MIR" /rpt_type=dispersed
repeat_region	2616. .2916 /rpt_family="Alu" /rpt_type=dispersed	repeat_region	complement(18001. .18310) /rpt_family="Alu" /rpt_type=dispersed
misc_feature	3309. .3532 /note="similar to EST H12297 (EST cluster Hs.31147)"	repeat_region	complement(18420. .18591) /rpt_family="MIR" /rpt_type=dispersed
misc_feature	3865. .4028 /note="similar to EST H12297 (EST cluster Hs.31147)"	repeat_region	18732. .18767 /rpt_family="Simple_repeat" /rpt_type=tandem
repeat_region	4099. .4406 /rpt_family="Alu" /rpt_type=dispersed	CDS	18967. .20202 /gene="CHST5" /note="I-GlcNAc6ST; carbohydrate sulfotransferase 5; expressed in the small intestine and colon"
repeat_region	4793. .5250 /rpt_family="L2" /rpt_type=dispersed		/codon_start=1 /product="intestinal N-acetylglucosamine-6-O-sulfotransferase"
repeat_region	complement(5251. .5329) /rpt_family="MIR" /rpt_type=dispersed		/protein_id="AAG26326.1" /db_xref="GI:11023148"
repeat_region	complement(5768. .6080) /rpt_family="Alu" /rpt_type=dispersed		/translation="MGMRAVPKVAHSTRRPDAARMWLPFRSSKTVTVLLAQTCLL LFIISRPGSSPAGEDRVHVLVLSSWRSGSFLGQLFSQHPDVFYLMERAWHVWTTL SQGSAATLHMAVRDLMRISIFLCMDVDFDAYMPOSRNLSAFFNWATSRALCSPACSAF PRGTISKQDVCKTLCTRQPFSLAREACRSYSHVVLKEVRFNLQVLYPLLSDPALNLR IVHLVRDPRAVLRSREAAGPILARDNGIVLGTNGKWEADPHLRLIREVCRSHVRIAE AATLKPPFPLRGYRLVRFEDLAREPLAEIRALYAFGLTLPQLEAWIHNIHSGSI GKPIEAFHTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQLLGYRPVYSADQORDL TLDLVLPRGPDHFSWASPD"
repeat_region	6111. .6567 /note="similar to EST AA001593 (EST cluster Hs.6853)"	repeat_region	20504. .20806 /rpt_family="Alu" /rpt_type=dispersed
repeat_region	8035. .8350 /rpt_family="Alu" /rpt_type=dispersed	repeat_region	20869. .20961 /rpt_family="L1" /rpt_type=dispersed
misc_feature	10545. .11049 /note="similar to EST AI950923 (EST cluster Hs.6853)"	repeat_region	20962. .21200 /rpt_family="MER4-group" /rpt_type=dispersed
repeat_region	11043. .11195 /rpt_family="Alu" /rpt_type=dispersed	repeat_region	21325. .21429 /rpt_family="MER4-group" /rpt_type=dispersed
repeat_region	11200. .11222 /rpt_family="Simple_repeat" /rpt_type=tandem	repeat_region	21430. .21722 /rpt_family="Alu" /rpt_type=dispersed
repeat_region	11570. .11871 /rpt_family="Alu" /rpt_type=dispersed	repeat_region	21725. .22007 /rpt_family="Alu" /rpt_type=dispersed
repeat_region	complement(12065. .12198) /rpt_family="L1" /rpt_type=dispersed	repeat_region	22008. .22040 /rpt_family="Simple_repeat" /rpt_type=tandem
repeat_region	complement(12205. .12493) /rpt_family="Alu" /rpt_type=dispersed	repeat_region	22041. .22313 /rpt_family="MER4-group" /rpt_type=dispersed
repeat_region	complement(12813. .13110) /rpt_family="Alu" /rpt_type=dispersed	repeat_region	22325. .22593 /rpt_family="Alu" /rpt_type=dispersed
mRNA	join(14185. .14228,16169. .16263,17712. .20821) /gene="CHST5" /product="intestinal N-acetylglucosamine 6-O-sulfotransferase"	repeat_region	22594. .22949 /rpt_family="MER4-group" /rpt_type=dispersed
gene	14185. .20821 /gene="CHST5" 14681. .14713 /rpt_family="Simple_repeat"	repeat_region	23116. .23221 /rpt_family="Simple_repeat" /rpt_type=tandem
repeat_region	complement(14715. .14990) /rpt_family="Alu" /rpt_type=dispersed	repeat_region	23344. .23400 /rpt_family="MaLR" /rpt_type=dispersed
repeat_region	complement(14995. .15288) /rpt_family="Alu" /rpt_type=dispersed	repeat_region	23433. .23456
repeat_region	complement(15292. .15589) /rpt_family="Alu"		

Fri Feb 1 20:20:49 2002

RESULT 6
AC009163
LOCUS
DEFINITION
AC009163 179065 bp DNA HTG 04-MAY-2000
Homo sapiens chromosome 16 clone RP11-77K12, WORKING DRAFT
SEQUENCE, 51 unordered pieces.
AC009163
AC009163.4 GI:7689976
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 179065)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
Unpublished
2 (bases 1 to 179065)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On May 4, 2000 this sequence version replaced gi:7239429.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 436185
Center clone name: RPCI-11_77K12

Summary Statistics
Consensus quality: 132269 bases at least Q40
Consensus quality: 157541 bases at least Q30
Consensus quality: 162978 bases at least Q20
Estimated insert size: 63060; agarose-fp estimation
Estimated insert size: 174065; sum-of-contigs estimation
Quality coverage: 3.87 in Q20 bases; agarose-fp estimation
Quality coverage: 1.4 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 51 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1020: contig of 1020 bp in length
1021 1120: gap of unknown length
1121 2488: contig of 1368 bp in length
2489 2588: gap of unknown length
2589 4067: contig of 1479 bp in length
4068 4167: gap of unknown length
4168 5211: contig of 1044 bp in length
5212 5312: contig of 1105 bp in length
5312 6417: gap of unknown length
6417 6517: contig of 1371 bp in length
6517 7888: gap of unknown length
7888 9089: contig of 1101 bp in length
9089 9189: gap of unknown length
9189 10212: contig of 1023 bp in length
10212 10312: gap of unknown length
10312 11714: contig of 1402 bp in length
11714 11814: gap of unknown length
11814 13074: contig of 1260 bp in length
13074 13174: gap of unknown length
13174 14854: contig of 1681 bp in length
14854 14955: gap of unknown length
14955 16910: contig of 1955 bp in length
16910 17010: gap of unknown length
17010 18668: contig of 1659 bp in length
18668 18768: gap of unknown length
18768 18669

18769
20274
20374
21427
21527
22911
23011
25171
25172
25272
27707
27807
29084
29184
31234
31334
31334
33855
33955
36108
36208
38157
38257
39988
40088
41610
41710
43401
43501
45246
45346
47680
47780
49624
49724
51989
52089
54401
54501
57338
57438
59991
60091
62555
62655
65582
65682
70062
70162
74923
75023
79849
79949
83053
83153
87069
87169
93513
93613
98371
98471
102252
102352
107788
107888
112455
112555
120361
120461
126855
126955
134019
134119
146224
146324

20273: contig of 1505 bp in length
20373: gap of unknown length
21426: contig of 1053 bp in length
21526: gap of unknown length
22910: contig of 1384 bp in length
23010: gap of unknown length
25171: contig of 2161 bp in length
25271: gap of unknown length
27706: contig of 2435 bp in length
27806: gap of unknown length
29083: contig of 1277 bp in length
29183: gap of unknown length
31233: contig of 2050 bp in length
31333: gap of unknown length
33854: contig of 2521 bp in length
33954: gap of unknown length
36107: contig of 2153 bp in length
36207: gap of unknown length
38156: contig of 1949 bp in length
38256: gap of unknown length
39987: contig of 1731 bp in length
40087: gap of unknown length
41609: contig of 1522 bp in length
41709: gap of unknown length
43400: contig of 1691 bp in length
43500: gap of unknown length
45245: contig of 1745 bp in length
45345: gap of unknown length
47679: contig of 2334 bp in length
47779: gap of unknown length
49623: contig of 1844 bp in length
49723: gap of unknown length
51988: contig of 2265 bp in length
52088: gap of unknown length
54400: contig of 2312 bp in length
54500: gap of unknown length
57337: contig of 2837 bp in length
57437: gap of unknown length
59990: contig of 2553 bp in length
60090: gap of unknown length
62554: contig of 2464 bp in length
62654: gap of unknown length
65581: contig of 2927 bp in length
65681: gap of unknown length
70061: contig of 4380 bp in length
70161: gap of unknown length
74922: contig of 4761 bp in length
75022: gap of unknown length
79848: contig of 4826 bp in length
79948: gap of unknown length
83052: contig of 3104 bp in length
83152: gap of unknown length
87068: contig of 3916 bp in length
87168: gap of unknown length
93512: contig of 6344 bp in length
93612: gap of unknown length
98370: contig of 4758 bp in length
98470: gap of unknown length
102251: contig of 3781 bp in length
102351: gap of unknown length
107787: contig of 5436 bp in length
107887: gap of unknown length
112454: contig of 4567 bp in length
112554: gap of unknown length
120360: contig of 7806 bp in length
120460: gap of unknown length
126854: contig of 6394 bp in length
126954: gap of unknown length
134018: contig of 7064 bp in length
134118: gap of unknown length
146223: contig of 12105 bp in length
146323: gap of unknown length
157038: contig of 10715 bp in length

(I-GlcNac-6-ST) gene, complete cds.

AF176839
AF176839.1 GI:5917707

human.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 1462)
Lee,J.K., Bhakta,S., Rosen,S.D. and Hemmerich,S.
Cloning and characterization of a mammalian
N-acetylglucosamine-6-sulfotransferase that is highly restricted to
intestinal tissue

JOURNAL
Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)

MEDLINE
99423499

REFERENCE
2 (bases 1 to 1462)
Lee,J.-K., Bhakta,S., Rosen,S.D. and Hemmerich,S.
Direct Submission
Submitted (10-AUG-1999) Respiratory Diseases, Roche Bioscience,
3401 Hillview Avenue, Palo Alto, CA 94304, USA

JOURNAL
Location/Qualifiers

FEATURES
1..1462
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q23.1"
/tissue_type="intestine"
<229..>1401
/gene="I-GlcNac-6-ST"
/product="N-acetylglucosamine 6-O-sulfotransferase"
<229..>1401
/gene="I-GlcNac-6-ST"
229..1401
/gene="I-GlcNac-6-ST"
/note="possibly involved in biosynthesis of mucosal
ligands for L-selectin"
/codon_start=1
/product="N-acetylglucosamine 6-O-sulfotransferase"
/protein_id="AAD56001.1"
/db_xref="GI:5917708"
/translation="MWLPFRFSKTVTVLLLAQTTCLLLEIISRPGSPSPAGGEDRVHV
LVLSWRSQSSFLQLEFSQHPDVFLMEPAWHVWVTLISQSSAATLHMAVRDLMRSIFL
CDMDVFDAYMPOSRNLSAFENWATSRALCSPACSAPFRGTISKQDVCKTLCTRQPS
LAREACRSYSHVVLKEVFENLVLYPLLSDPALNLRIVHLVDRPRAVLRREAAGPI
LARDNGIVLGTNGKWEADPHLRILREVCRSVRIAEATLKPFPFLRGYRLVREFD
LAREPLAEIRALYAFGLTLPQLEAWIHNITHSGSIGCKPIEAFHTSSRNARNVSOAW
RHALPFTKILRVQEVFCALQLIGYRPVYSADQQRDLTLDLVLPRGPDHFSWASPD"

BASE COUNT 239 a 510 c 446 g 267 t

ORIGIN

Query Match 61.1%; Score 1248.6; DB 9; Length 1462;
Best local Similarity 99.7%; Pred. No. 3.3e-197;
Matches 1251; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 197 cgggtgccccagccgcgcgatgtggtgccacgtgttccacgaagacagtgacagtg 256
|||
Db 208 CGAAGGCCCCAGCCGCGCATGTGGTGCCACGGTCTCCAGCAAGACAGTGACAGTG 267
|||
QY 257 ctctcctggcacagaccactgcctcctgctcttcatcatctcccgccagggccctca 316
|||
Db 268 CTCCTCCTGGCACAGACCACCTGCCTCCTGCTCTTCATCATCATCTCCGGCCAGGGCCCTCA 327
|||
QY 317 tccccagccggcggcagagatcgtgtgcacgtgctggtgtgtgtcctcgtggtgcgtcgggc 376
|||
Db 328 TCCCCAGCCGGCGGAGGATCGTGTGCACGTGCTGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 387
|||
QY 377 tcataccttctggccagctcttccagccagcccccagcgtcttctacactgatggagccc 436
|||
Db 388 TCATCCTTCTTGGGCCAGCTCTTCAGCCAGCACCCCGACGCTCTTACCTGATGGAGCCC 447
|||
QY 437 gcgtggcatgtgtggaccaccctgtcgcagggcagcgcggcaacgtgcacatggccgtg 496
|||

448 GCGTGGCATGTGTGGACCACCOCTGTGGCAGGGGAGCGGGCAACGCTGCACATGGCCCGTG 507
QY 497 cgcgacctgatgcgctctatctttttgtggacatggacgtgtttgatgcctaatgccaa 556
|||
Db 508 CGCGACCTGTATGGGCTCTATCTTTTGTGGACATGGACGTGTTTGTATGCTATACATGCCA 567
|||
QY 557 cagagccgaaacctgtccgcctttttcaactgggcaacagagccgcgcgtgtgtgcgcg 616
|||
Db 568 CAGAGCCGAAACCTGTCCGCCCTTTTCAACTGGGCAACGAGCCGCGCTGTGTGTCGCCG 627
|||
QY 617 cccgcctgcagcgcctttccccagggcacccatcagcaagcaggacgtatgcaagacactg 676
|||
Db 628 CCGGCTGCAGCGCCTTCCCGAGGCACCATCAGCAAGCAGGACGTATGCAAGACACTG 687
|||
QY 677 tgcacgcggcagcattcagccttgcccggggagcctgcgcgtcctacagccacgtgtgtg 736
|||
Db 688 TGCACGGCGCAGCCATTTCAGCCTGGCCCGGAGGCCCTTCCCGCTCCTACAGCCACGTGGTG 747
|||
QY 737 ctcaagaggaggtgcgcttcttcaacctgcaggtgctctacccgctgctcagcgcaccccg 796
|||
Db 748 CTCAAGGAGGTGGCTTCTTCAACCTGCAGGTGCTCTACCCGCTGCTCAGCGACCCCGCG 807
|||
QY 797 ctcaacctgcgcctgctgcacctggtgcgcgaccccgcgcgccgtgctgcgtcccgagg 856
|||
Db 808 CTCAACCTGCGCATCGTGCACCTGCTGCGGACCCCGCGGCGCTGCTGCGCTCCCGGGAG 867
|||
QY 857 gcgcggggcccgatactgacgcgcacaacggcatcgtgctgggcaaccaacggcaagtgg 916
|||
Db 868 GCGCGGGCCCGGATCTGGCACCGGCAACGGCATCGTGTGGGACCAACGCGCAAGTGG 927
|||
QY 917 gtggaggccgacccctcacctgcgcctgattgcgcgaggtgtgcccgcgcgtaccgcctg 976
|||
Db 928 GTGGAGGCGGACCCCTCACCTGCGCCTGATTCGGGAGGTGTGCCGCGACGACGTCGCATC 987
|||
QY 977 gccgaggccgcacactcaagccgccaccttctctcgcgcgcgtaccgcctggtgcgc 1036
|||
Db 988 GCCGAGGCCCGCACACTCAAGCGCGCCACCCCTTCTTCCGCGGCGCTACCGCTGCTGCGC 1047
|||
QY 1037 ttgcaggacctggcgggagccgctgagcgcgtggcagatccgcgcactctacgccttccac 1096
|||
Db 1048 TTCGAGGACCTGGCGGGGAGCGCTGGCAGAGATCCGCGCACTCTACGCCCTTCACCCGGC 1107
|||
QY 1097 ctgacctcaccgcacagctcagggcctggatccacaacatcacccacgggtcggggatc 1156
|||
Db 1108 CTGACCCCTACGCCACAGCTCGAGGCTGGATCCACACATCACCCACGCGGTGGGGATC 1167
|||
QY 1157 ggcaagcccaatcgagccctccatacttctgcttaggaatgcgcgaacgtctcccagggc 1216
|||
Db 1168 GGCAAGCCAATCGAGGCCCTTCCATACTTCGTCTAGGAATGCGCGCAACGCTCTCCAGGCC 1227
|||
QY 1217 tggcgccacgcgttgcccttcaactaagatcctgcgcgtgcaggaggtgtgcgcgcggcg 1276
|||
Db 1228 TGGCGCCACGCGTGGCCCTTCACTAAGATCCTGCGCGTGGCAGAGGTGTGCGCGCGCGCG 1287
|||
QY 1277 ctgcagctgctgggctaccgctgtgtgtactctgcggaccagcagcgtgacctcaccctg 1336
|||
Db 1288 CTGCAGCTGCTGGGCTACCGGCGCTGTGTACTCTGCGGACCCAGCAGCGTGACCTCACCCCTG 1347
|||
QY 1337 gatctggtgctgccacgagggcccagaccacttcagctgggcatgcctgactgagaaactc 1396
|||
Db 1348 GATCTGCTGCTGCCACGAGGCGCCAGACCACTTCAGTGGGCAATCGCCTGACTGAGAACTC 1407
|||
QY 1397 tgggccttagagcaagccccgaactgtggtgcgcagggccccaggaagcagcactgcat 1451
|||
Db 1408 TGGGCCCTTAGAGCAGGCCCGGAACTGTGCTGCCAGGCCCGCCAGGAGGCGACTGCAT 1462
|||

RESULT 8
AF280086 3786 bp mRNA PRI 20-FEB-2001
LOCUS Homo sapiens N-acetylglucosamine 6-O-sulfotransferase GST-4beta
DEFINITION mRNA, complete cds.
ACCESSION AF280086

QY 1566 ttttcctttgagttcttctggagctgcttctctcatcagtgctcactcttcatggaagcaa 1625
|||||
Db 136379 TTTTCTTTGAGTCCTCTGGAGCTGCCTTCTCATCAGGTGCATCTTCATGG-AAAGCAA 136321
|||||
QY 1626 ctcttgccctacacctcttctgtggcgagggaggaatt 1663
|||||
Db 136320 CTCTTGCCCCCT-CCTCCTCTGGGCACAGGGTGTGCGTT 136284
|||||

RESULT 11
AC009105 208185 bp DNA HTG 25-APR-2001
LOCUS Homo sapiens chromosome 16 clone RP11-455E15, WORKING DRAFT
DEFINITION SEQUENCE, 17 unordered pieces.
AC009105
AC009105.7 GI:13786304
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 208185)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
Unpublished
2 (bases 1 to 208185)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 25, 2001 this sequence version replaced gi:7689928.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 0
Center clone name: RPCI-11_455E15

Summary Statistics
Consensus quality: 180040 bases at least Q40
Consensus quality: 192095 bases at least Q30
Consensus quality: 195726 bases at least Q20
Estimated insert size: 195650; agarose-fp estimation
Estimated insert size: 206585; sum-of-contigs estimation
Quality coverage: 6.57 in Q20 bases; agarose-fp estimation
Quality coverage: 6.23 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1062: contig of 1062 bp in length
* 1063 1162: gap of unknown length
* 1163 2359: contig of 1197 bp in length
* 2360 2459: gap of unknown length
* 2460 3653: contig of 1194 bp in length
* 3654 3753: gap of unknown length
* 3754 5640: contig of 1887 bp in length
* 5641 5740: gap of unknown length
* 5741 7370: contig of 1630 bp in length
* 7371 7470: gap of unknown length
* 7471 8584: contig of 1114 bp in length
* 8585 8685: gap of unknown length
* 8685 10014: contig of 1330 bp in length
* 10015 10114: gap of unknown length
* 10115 12451: contig of 2337 bp in length
* 12452 12551: gap of unknown length
* 12552 28715: contig of 16164 bp in length

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 18:15:53 ; Search time 315.85 Seconds
(without alignments)
5548.117 Million cell updates/sec

Title: US-09-593-828-3
Perfect score: 2044
Sequence: 1 ggctcgaggtccactgtgct.....aaaaaaaaaaaaaaaaaaaa 2044

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues 1861242

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101:*

1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2044	100.0	2044	22	AAD02699 Human glycosyl sul
2	1973	96.5	2170	22	AAD02698 Human glycosyl sul
3	1827.2	89.4	2988	21	AAC76156 Human ORFX ORF1711
4	1753.6	85.8	160552	22	AAD02697 Human glycosyl sul
5	1111.4	54.4	1694	22	AAD02700 Human glycosyl sul
6	751.2	36.8	1989	22	AAD02696 Mouse glycosyl sul
7	400	19.6	2032	20	AAZ20792 Human glycosyl sul
8	377	18.4	2065	21	AAZ94211 Human transferase
9	354	17.3	1926	20	AAZ20793 Mouse glycosyl sul
10	235.2	11.5	2156	19	AAV21200 Glycosaminoglycan
11	211	10.3	1043	21	AAA14459 Human interleukin-

12	211	10.3	1601	21	AAF21313 Human low adenosin
13	211	10.3	1601	21	AAA35191 Human adenosine re
14	211	10.3	1601	21	AAZ46814 Interleukin-10 (IL
15	211	10.3	1618	12	AAQ10207 pH15C insert conta
16	211	10.3	1618	14	AAQ46958 Human cytokine syn
17	211	10.3	1618	21	AAZ86905 Human CSIF coding
18	211	10.3	1618	22	AAF90433 Human cytokine syn
19	211	10.3	1618	22	AAF82588 Human CSIF cDNA.
20	211	10.3	1618	22	AAF83186 Human interleukin-
21	211	10.3	1645	21	AAZ58659 Human cDNA sequenc
22	211	10.3	2797	22	AAH18187 Human low adenosin
23	211	10.3	15630	21	AAF21317 Human adenosine re
24	211	10.3	15630	21	AAA35195 Human tissue facto
25	208.2	10.2	13865	19	AAV40401 Human cholesteryl
26	208.2	10.2	13865	20	AAZ32165 Human thymopoietin
27	207.4	10.1	3234	16	AAQ92781 Human inflammatory
28	205.8	10.1	700	22	AAH92085 TXA2 receptor gene
29	205.8	10.1	2932	13	AAQ25388 Human thromboxane
30	205.8	10.1	2932	20	AAZ32161 Human endothelial
31	205.8	10.1	2932	20	AAZ32162 SHOX gene prelimin
32	205.6	10.1	15577	19	AAV35616 Human SHOX (short
33	205.6	10.1	32367	19	AAV35620 Human calcium sens
34	205.4	10.0	48000	22	AAF27996 DNA sequence of th
35	205.2	10.0	54548	21	AAZ45596 Human kidney relat
36	204.8	10.0	9456	22	AAI63326 Insulin receptor g
37	204.6	10.0	7240	22	AAH50570 Human insulin rece
38	204.6	10.0	7240	22	AAH31170 Human insulin rece
39	204.6	10.0	7240	22	AAH31267 Human insulin rece
40	204.6	10.0	7240	22	AAD04467 Human insulin rece
41	204.4	10.0	84607	20	AAZ90847 Human PACAP genom
42	204.2	10.0	72928	20	AAZ18355 Human ASTH1J 5' ge
43	204.2	10.0	72928	21	AAZ80253 Human ASTH1J 5' ge
44	204	10.0	119950	20	AAZ90201 Human yes1 gene.
45	203.4	10.0	1227	22	AAF45078 Human secreted pro

ALIGNMENTS

RESULT 1	
AAD02699	
ID	AAD02699 standard; cDNA; 2044 BP.
XX	
AC	AAD02699;
XX	
DT	02-MAY-2001 (first entry)
XX	
DE	Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #2.
XX	
KW	Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation;
KW	systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW	polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW	glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenailitis;
KW	Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW	demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW	myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW	asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW	chromosome 16q23.1; ss.
XX	
OS	Homo sapiens.
XX	
FT	Key Location/Qualifiers
FT	CDS 218..1390
FT	/*tag= a
FT	/product= "Human glycosyl sulfotransferase-4alpha
FT	(GST-4alpha)"
FT	/note= "CDS is specifically claimed as SEQ ID NO: 4
FT	in claim 6 (page no: 41) of the specification"
XX	
PN	WO200106015-A1.
XX	
PD	25-JAN-2001.

Db 1501 ||||| gggagtttggggtcctccctgaagtaagcaaggactgaagctttcttctctctgatt 1560
QY 1561 ctcgggtttctcttttgagtcttctggagctgccttctcatcagggtgcactcttcatgaaa 1620
Db 1561 ||||| ctcgggtttctcttttgagtcttctggagctgccttctcatcagggtgcactcttcatgaaa 1620
QY 1621 agcaactttgcccctacctcttctgggagggagtaagttactgtctaaattaaatta 1680
Db 1621 ||||| agcaactttgcccctacctcttctgggagggagtaagttactgtctaaattaaatta 1680
QY 1681 aatgtgtgccaggccgggtgctctcatgcctgttaatcccagcattttgagaggctg 1740
Db 1681 ||||| aatgtgtgccaggccgggtgctctcatgcctgttaatcccagcattttgagaggctg 1740
QY 1741 aggcgggtggatcacctgaggtcaggattcaaaaaccagcctggccaacatagtgaaacc 1800
Db 1741 ||||| aggcgggtggatcacctgaggtcaggattcaaaaaccagcctggccaacatagtgaaacc 1800
QY 1801 cctcttactaaaaatgcaaaaattagtcggcggtggcgacacactcctgtaatcccagc 1860
Db 1801 ||||| cctcttactaaaaatgcaaaaattagtcggcggtggcgacacactcctgtaatcccagc 1860
QY 1861 tacttaggaggtgaggtgggaaaaatcacttgactccaaagggtggaggttcagtaagc 1920
Db 1861 ||||| tacttaggaggtgaggtgggaaaaatcacttgactccaaagggtggaggttcagtaagc 1920
QY 1921 tgaatatcatgcaactgcaccctagcttgggtggcaagcaaaactctataaaaaataa 1980
Db 1921 ||||| tgaatatcatgcaactgcaccctagcttgggtggcaagcaaaactctataaaaaataa 1980
QY 1981 ttaataattgttcaaaagtccctgcgaaaaaaataaaaaaaataaaaaaaataa 2040
Db 1981 ||||| ttaataattgttcaaaagtccctgcgaaaaaaataaaaaaaataaaaaaaataa 2040
QY 2041 aaaa 2044
Db 2041 |||| aaaa 2044

RESULT 2
AAD02698
ID AAD02698 standard; cDNA; 2170 BP.
XX
AC AAD02698;
XX
DT 02-MAY-2001 (first entry)
XX
DE Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #1.
XX
KW Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;
KW therapy; selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
XX chromosome 16q23.1; ss.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 5'UTR 9..343 /*tag= a
FT CDS 344..1516 /*tag= b
FT FT /product= "Human glycosyl sulfotransferase-4alpha
FT (GST-4alpha)"
FT /note= "CDS is specifically claimed as SEQ ID NO: 4
FT in claim 6 (page no: 41) of the specification"
FT 3'UTR 1517..2134

FT XX
PN WO200106015-A1.
XX 25-JAN-2001.
XX 19-JUL-2000; 2000WO-US19741.
XX 20-JUL-1999; 99US-0144694.
PR 13-JUL-2000; 2000US-0593828.
XX (REGC) UNIV CALIFORNIA.
PA Rosen SD, Lee JK, Hemmerich S;
XX WPI; 2001-138471/14.
DR P-PSDB; AAY72639.
XX
PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT diagnostic and therapeutic agent screening applications -
XX
PS Claim 6; Page 62; 128pp; English.
XX
CC The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
CC alpha) cDNA. GST-4 gene is found on chromosome 16q23.1.
CC GST is a type 2 membrane protein useful for inhibiting a binding event
CC between a selectin and a selectin ligand, which comprises contacting the
CC selectin with a non-sulphated selectin ligand, GST and a small molecular
CC agent that inhibits the sulphation activity of GST. GST is also useful
CC in inhibiting a selectin mediated binding event. GST is useful in gene
CC therapy to treat disorders such as acute or chronic inflammation,
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation.
XX
SQ Sequence 2170 BP; 449 A; 682 C; 612 G; 427 T; 0 other;

Query Match 96.5%; Score 1973; DB 22; Length 2170;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1998; Conservative 0; Mismatches 5; Indels 2; Gaps 2;
QY 41 gaaagctcccagtgaggaaactgttcttctggagactctgtgtggcatagagtattcaac 100
Db ||||| gaaagctcccagtgaggaaactgttcttctggagactctgtgtggcatagagtattcaac 100
QY 101 caccttaagaagacctctgttcttctggaaacacagatgtcgagacatctcccatgatt 160
Db ||||| caccttaagaagacctctgttcttctggaaacacagatgtcgagacatctcccatgatt 160
QY 161 tgtgatcagcgttgacgtctcccagcagccctgacggtggccccagccgcccgcattg 220
Db ||||| tgtgatcagcgttgacgtctcccagcagccctgacggtggccccagccgcccgcattg 220
QY 221 tggctgccacggttctccagcaagacagtgacagtgtcctcctctggcacagaccacctgc 280
Db ||||| tggctgccacggttctccagcaagacagtgacagtgtcctcctctggcacagaccacctgc 280
QY 281 ctctgtcttctcatcatctcccggccagggccctcatccccagccggcgaggatcgt 340
Db ||||| ctctgtcttctcatcatctcccggccagggccctcatccccagccggcgaggatcgt 340
QY 341 gtgcacgtgtgtgtctctctgtgtggcgtcgtgggtcattcttcttggccagctcttc 400
Db ||||| gtgcacgtgtgtgtctctctgtgtggcgtcgtgggtcattcttcttggccagctcttc 400
QY 401 agccagacccccgacgtcttctacctgatggagcccggtggcatgtgtggaccacctg 460


```

PR 02-APR-1999;    99US-0127636.
PR 05-APR-1999;    99US-0127728.
PR 30-MAR-2000;    2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach M;
PI
XX
DR WPI; 2000-602362/57.
DR P-PSDB; AAB41947.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 5; Page 2597-2599; 5507pp; English.
PS
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulneryary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; antifungal; antirheumatic;
CC antiinflammatory; antibacterial; antiviral; antifungal; immunosuppressive;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 2988 BP; 613 A; 916 C; 842 G; 617 T; 0 other;

Query Match      89.4%; Score 1827.2; DB 21; Length 2988;
Best Local Similarity 95.4%; Pred. No. 1.6e-292;
Matches 1915; Conservative 0; Mismatches 88; Indels 5; Gaps 3

QY 41 gaaagtccacgtgaggaaactgttcttcttggaactctgtggcatagagtgcattcaac 100
Db |||||
QY 982 gaaagtccccagtgaggaaactgttcttcttggaactctgtggcatagagtgcattcaac 1041
Db |||||
QY 101 caccttaagaagacctctgtgctttccttggaacacagatgtogagacatctcccattgatt 160
Db |||||
QY 1042 caccttaagaagacctctgtgctttccttggaacacagatgtogagacatctcccattgatt 1101
Db |||||
QY 161 tgtgatcagcgttgacgtctctccagcagccctggacgtgtgcccccagcgcccgcatg 220
Db |||||
QY 1102 tgtgatcagcgttgacgtctctccagcagccctggacgtgtgcccccagcgcccgcatg 1161
Db |||||
QY 221 tggctgccacggttctccagcaagacagtagcacagtgtctctccttgccacagaccactgc 280
Db |||||
QY 1162 tggctgccacggttctccagcaagacagtagcacagtgtctctccttgccacagaccactgc 1221
Db |||||
QY 281 ctctgtcttcttcattctcccgccagggccctcatccccagcgccggcgaggatcgt 340
Db |||||
QY 1222 ctctgtcttcttcattctcccgccagggccctcatccccagcgccggcgaggatcgt 1281
Db |||||
QY 341 gtgcacgtgctggtgctctcctcgtggcgctcgggtcattcttcttgggccaagctcttc 400
Db |||||
QY 1282 gtgcacgtgctggtgctctcctcgtggcgctcgggtcattcttcttgggccaagctcttc 1341
Db |||||
QY 401 agccagcaccgccgactcttctacactgatggagcccgcggtggcatgtgtggaccacctg 460
Db |||||
QY 1342 agccagcaccgccgactcttctacactgatggagcccgcggtggcatgtgtggaccacctg 1401
Db |||||

```


Fri Feb 1 20:20:50 2002

QY 1537 tgcacgtttctctctcctgattctcgtgtttctccttttgagttctcttgagagtcctttct 1596
Db 2482 tgcacgtttctctctcctgattctcgtgtttctccttttgagttctcttgagagtcctttct 2541
QY 1597 catcaggtgcactcttcattggaagcaactcttgccctacctctcttgaggcgagggga 1656
Db 2542 catcaggtgcactcttcattggaagcaactcttgccctacctctcttgaggcgagggga 2600
QY 1657 gtaagttactgctaaattaaattaaatgtgtgccaggccgggtgcggtggtcctcatgcctg 1716
Db 2601 gtaagttactgctaaattaaattaaatgtgtgccaggccgggtgcggtggtcctcatgcctg 2660
QY 1717 taatccagcattttgagaggtgagcggtggatcacctgaggtcaggattcaaaacc 1776
Db 2661 taatccagcattttgagaggtgagcggtggatcacctgaggtcaggattcaaaacc 2720
QY 1777 agcctggccaacatagtgaaacccctctctactactaaaaatgcaaaaaattagtcggcgctg 1836
Db 2721 agcctggccaacatagtgaaacccctctctactactaaaaatgcaaaaaattagtcggcgctg 2780
QY 1837 gtggcacactcctgtaatccagctactactagggctgaggtgggaaatcacttggact 1896
Db 2781 gtggcacactcctgtaatccagctactactagggctgaggtgggaaatcacttggact 2840
QY 1897 ccaaaggtggaggtgcagtaagctgaaatcaatgcactgcacccctagcttggtgggcaa 1956
Db 2841 ccaaaggtggaggtgcagtaagctgaaatcaatgcactgcacccctagcttggtgggcaa 2900
QY 1957 agcaaaaactctatcaaaaaataattaataattgttcaaaagtcccgcaaaaaaaa 2016
Db 2901 agcaaaaactctatcaaaaaataattaataattgttcaaaagtcccgcaaaaaaaa 2960
QY 2017 aaaaaaaaataaaaaataaaaaa 2044
Db 2961 aaaaaaaaataaaaaataaaaaa 2988
RESULT 4
AAD02697
ID AAD02697 standard; DNA; 160552 BP.
XX
AC AAD02697;
XX
DT 02-MAY-2001 (first entry)
XX Human glycosyl sulfotransferase-4 (GST-4) genomic DNA.
DE Human; glycosyl sulfotransferase-4; GST-4; immunosuppressive;
KW therapy; selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW chromosome 16q23.1; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT exon 32847..32922
FT /*tag= a
FT /number= 1
FT /label= 4a_5U4
FT 32923..35592
FT /*tag= b
FT /cons_splice= (5'site:NO, 3'site:YES)
FT 35593..35674
FT /*tag= c
FT /number= 2
FT /label= 4a_5U3
FT 35675..45093
FT intron

FT exon /*tag= d
FT 45094..45185
FT /*tag= e
FT /number= 3
FT /label= 4a_5U2
FT 45186..46633
FT /*tag= f
FT /cons_splice= (5'site:NO, 3'site:NO)
FT 46634..46700
FT /*tag= g
FT /number= 4
FT /label= 4a_5U1
FT 46701..47938
FT /*tag= h
FT /cons_splice= (5'site:YES, 3'site:NO)
FT 47939..49746
FT /*tag= i
FT /number= 5
FT /note= "Includes 17 base pairs of 5'UTR, the ORF
FT and all of 3'UTR"
FT 47939..47955
FT /*tag= j
FT /note= "Portion of 5' untranslated region (5'UTR)"
FT 47956..49128
FT /*tag= k
FT /product= "Human glycosyl transferase-4alpha
FT (GST-4alpha)"
FT 49129..49746
FT /*tag= l
FT 83257..83347
FT /*tag= m
FT /label= 4a_5U2
FT 83348..96412
FT /*tag= n
FT /cons_splice= (5'site:NO, 3'site:NO)
FT 96413..96484
FT /*tag= o
FT /label= 4a_5U1
FT 96485..98456
FT /*tag= p
FT /cons_splice= (5'site:NO, 3'site:NO)
FT 98457..99968
FT /*tag= q
FT /note= "Includes 17 base pairs of 5'UTR, the ORF
FT and all of 3'UTR"
FT 98457..98473
FT /*tag= r
FT /note= "Portion of 5' untranslated region (5'UTR)"
FT 98474..99661
FT /*tag= s
FT /product= "Human glycosyl transferase-4beta
FT (GST-4beta)"
FT 99662..99968
FT /*tag= t
XX WO200106015-A1.
PN 25-JAN-2001.
XX
XX 19-JUL-2000; 2000WO-US19741.
PF
XX
XX 20-JUL-1999; 99US-0144694.
PR 13-JUL-2000; 2000US-0593828.
XX (REGC) UNIV CALIFORNIA.
XX Rosen SD, Lee JK, Hemmerich S;
PI WPI; 2001-138471/14.
XX P-PSDB; AAY72639, AAY72640.
DR New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT diagnostic and therapeutic agent screening applications -

Db 948 gagggccacactcaagccgcccacccctttctgcgccgctaccgctggtgcgcttc 1007
QY 1040 gaggacctggcgccgagccgctgagcagatccgcgactctacgcttccaccggcctg 1099
Db 1008 gaggacctggcgccgagccgctgagcagatccgcgactctacgcttccaccggcctg 1067
QY 1100 accctcagccacagctcgaggccctgagctcagcacaacatcacccaggggtcgggatcgcc 1159
Db 1068 agtctcagccacagctcgaggccctgagctcagcacaacatcacccagggatcgacccgtgt 1127
QY 1160 aagccaatcgaggccctcactctgctaggaatcgcgcaacgtctccagggcctgg 1219
Db 1128 gcgcgcccgaagccttcaagactctgctcaggaatcgctcaacgtctccagggcctgg 1187
QY 1220 cgcacagcgttgcccttcaactaagatccctgcgcgtgcaggaggtgtgcgcggcgccgtg 1279
Db 1188 cgccatgcgtgcccttgcacaagatccgcgcgtgcaggaggtgtgcgcggcgccgtg 1247
QY 1280 cagctgctgggtaccgctgtgtactctgcggaccagcagcgtgacctcaccctggat 1339
Db 1248 cagctgctgggtaccgctgtgtactctgcggaccagcagcgtgacctcaccctggat 1307
QY 1340 ctggtgctgccacagggccagaccacttccagctgggcatc--gcctgactgagaactct 1397
Db 1308 ctggtgctgccacagggcctgaacggcttccacttgggcatcaccacccgctcgacccc 1367
QY 1398 gggccttagagcaagcccc--gaactgtgtgctgcccagggccaggaagcagctgcatggtg 1455
Db 1368 cgaattagtgaggccacagctgttagcaggcggtagggccggaggagagtgcatggtg 1427
QY 1456 gaaaaggagctggcgccatgggaa--caggtccctactatcaacccgggagtttggggt- 1513
Db 1428 cagaggggctggggcgccagcaggaagcaggtccctatattgaccaaggagtttggtgta 1487
QY 1514 -----cctcccctgaagtaagcaaggactgcacgtttcttcttcttctctgattctcg 1564
Db 1488 cgacccctcccctcccacagtaggcaaggactgcacgtttcttcttcttctctgattctcg 1547
QY 1565 gtttctcttgagctcttctgagctgccttctcactcaggtgactcttctcaggaagca 1624
Db 1548 gtttctcttgagctcttctgagctgccttctcactcaggtgactcttctcaggaagca 1606
QY 1625 actcttgcccctacctcttctgagcgcagggaggaagtt 1663
Db 1607 actcttgcccctgctctcttctggcagcaggggtgctg 1645

RESULT 6
AAD02696
ID AAD02696 standard; cDNA; 1989 BP.
XX
AC AAD02696;
XX
DT 02-MAY-2001 (first entry)
XX
DE Mouse glycosyl sulfotransferase-4 (GST-4) cDNA.
XX
KW Mouse; glycosyl sulfotransferase-4; GST-4; immunosuppressive;
KW therapy; selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
XX chromosome 8E1; ss.
OS Mus musculus.
XX
FH
FT Key Location/Qualifiers
CDS 107..1294

FT
FT
FT
FT
XX
PN
XX
PD
XX
PF
XX
PR
PR
XX
PA
XX
PI
XX
DR
DR
XX
PT
XX
PS
XX
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
XX
SQ
Sequence 1989 BP; 415 A; 586 C; 550 G; 438 T; 0 other;
Query Match 36.8%; Score 751.2; DB 22; Length 1989;
Best Local Similarity 70.9%; Pred. No. 2e-115;
Matches 1046; Conservative 0; Mismatches 413; Indels 17; Gaps 3;
QY 202 gcccccagccgcccgcagctgtggtgcccacgggttctccagcaagacagtgacagtgctct 261
Db 91 ggccttagccagaggtatgcggtaccccggtttctccagcactgcatgtcttctgctct 150
QY 262 cctggcacagaccacctgctctctcttctcactatctcccgccagggccctcctccc 321
Db 151 gatggtacag--actggcactcctggtctctcctggtctcccgccaagtgccatcgctccc 207
QY 322 agccggcgccgaggtatcgtgtgcacgtgctggtgctctcctcgtggtcgctcggtcctc 381
Db 208 agcaggccttgggagcgtgtgcacgtgctggtgactgctcctcgtggtcggtcggtcgtc 267
QY 382 cttcttgggcccagctcttccagccagcaccgccagcttctcactgatggagcccgctg 441
Db 268 ctctgtggccagctcttccagccacaccccgatgtcttctcactgatggagcccgctg 327
QY 442 gcatgtgtggaccacccctgtcgcagggcgagcgcggcaacgctgcacatggccgtgcgca 501
Db 328 gcacgtctgggatacgttgtcgcagggcgagtgcccccgcaactccacatggccgtgcgtga 387
QY 502 cctgatgcgctctatctcttttctgacatggagcgtgtttgatgcctacatgccacagag 561

/*tag= a
/product= "Mouse glycosyl sulfotransferase-4 (GST-4)"
/note= "CDS is specifically claimed as SEQ ID NO: 2
in claim 6 (page no: 41) of the specification"
WO200106015-A1.
25-JAN-2001.
19-JUL-2000; 2000WO-US19741.
20-JUL-1999; 99US-0144694.
13-JUL-2000; 2000US-0593828.
(REGC) UNIV CALIFORNIA.
Rosen SD, Lee JK, Hemmerich S;
WPI; 2001-138471/14.
P-PSDB; AAY72638.
New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
diagnostic and therapeutic agent screening applications -
Claim 6; Fig 2; 128pp; English.
The present sequence is mouse glycosyl sulfotransferase-4 (GST-4) cDNA.
GST-4 gene is found on chromosome 8E1.
GST is a type 2 membrane protein useful for inhibiting a binding event
between a selectin and a selectin ligand, which comprises contacting the
selectin with a non-sulphated selectin ligand, GST and a small molecular
agent that inhibits the sulphation activity of GST. GST is also useful
in inhibiting a selectin mediated binding event. GST is useful in gene
therapy to treat disorders such as acute or chronic inflammation,
systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
dermatitis, myocarditis, regional enteritis, adult respiratory distress
syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
during transplantation.

XX
XX

Db 1153 aacaacctaaatttggttcttaggcgggcggggtggtcacgcctgtaatcccagcactt 1212

QY 1731 tgagagggctgagggcgggtgggtacacctgaggtcagga-ttcaaaaaccagcctggccaaca 1789

Db 1213 tgggagggctgagggcgggtgggtacacctgaggtcaggaagtccctaaccagcctgggtcaaca 1272

QY 1790 tagtgaaccocccctctctactaaaaatgcaaaaaattagtcggcgctggtggcacactcct 1849

Db 1273 tgggtgaaccocccgtctctactaaaaatataaaaaattagccgggcatggtggcgcgacct 1332

QY 1850 gtaatcccagctacttaggaggtgaggtgggaaaaatcaccttggactccaaaagggtggagg 1909

Db 1333 gtaatcccagctacttgggaggtgaggtgaggaagaaattgcttgaaccocccagagatggaag 1392

QY 1910 ttgcagtaagctgaaatcatgccactgcacccctagcttgggtggcaaaagcaaaaactctat 1969

Db 1393 ttgcagtgagctgatcatgccccctgtactccagcctgggtgacagagcaagactctgt 1452

QY 1970 caaaaaataataataaa 1988

Db 1453 ctcaaaaaataaaaaataa 1471

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:50:41 ; Search time 43.52 Seconds
(without alignments)
1310.804 Million cell updates/sec

Title: US-09-593-828-8
Perfect score: 2051
Sequence: 1 MWLPRFSSKTVTVLLLAQTT.....LTLDLVLPGRDPHFSWASPD 390

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues 473505

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2051	100.0	390	4	Q9UBY3	Q9uby3 homo sapien
2	2051	100.0	411	4	Q9GZS9	Q9gzs9 homo sapien
3	1729.5	84.3	395	4	Q9GZX3	Q9gzx3 homo sapien
4	1536.5	74.9	395	11	Q9QUP4	Q9gup4 mus sapien
5	1008	49.1	386	4	Q9Y5R3	Q9y5r3 homo sapien
6	965.5	47.1	388	11	Q9RLI1	Q9rlil mus musculu
7	958.5	46.7	388	11	Q9WUE5	Q9wue5 mus musculu
8	665	32.4	530	11	O88276	O88276 mus musculu
9	659.5	32.2	483	4	Q9GZN5	Q9gzn5 homo sapien
10	659.5	32.2	530	4	Q9Y6F2	Q9y6f2 homo sapien
11	656.5	32.0	484	4	Q9UED5	Q9ued5 homo sapien
12	656.5	32.0	531	4	Q9Y4C5	Q9y4c5 homo sapien
13	617.5	30.1	484	11	Q9EP78	Q9ep78 mus musculu
14	614.5	30.0	484	11	Q99NB0	Q99nb0 mus musculu
15	610	29.7	486	4	O75667	O75667 homo sapien
16	609	29.7	486	4	Q9NS84	Q9ns84 homo sapien
17	577.5	28.2	479	4	O75099	O75099 homo sapien
18	576	28.1	472	11	O88199	O88199 mus musculu
19	560	27.3	474	11	Q9QZL2	Q9qzl2 rattus norv

20	520	25.4	411	11	Q9EQC0	Q9eqc0 mus musculu
21	513.5	25.0	411	4	O43916	O43916 homo sapien
22	472.5	23.0	441	13	O93403	O93403 torpedo cal
23	184.5	9.0	183	5	Q9VMC2	Q9vmc2 drosophila
24	174.5	8.5	363	5	Q9VMC4	Q9vmc4 drosophila
25	151	7.4	114	11	Q9D0K5	Q9d0k5 mus musculu
26	107.5	5.2	315	5	Q9VMC3	Q9vmc3 drosophila
27	99.5	4.9	972	5	Q9VLX0	Q9vlix0 drosophila
28	95	4.6	474	3	O14192	O14192 schizosacch
29	94	4.6	582	2	O53639	O53639 mycobacteri
30	94	4.6	933	5	Q9GNZ1	Q9gnz1 leishmania
31	93.5	4.6	615	2	Q9RRS5	Q9rrs5 deinococcus
32	93	4.5	523	2	Q9I6Z6	Q9i6z6 pseudomonas
33	92	4.5	925	1	Q9HLE8	Q9hle8 thermoplasma
34	91	4.4	920	2	Q9A0U7	Q9a0u7 streptococc
35	89.5	4.4	1844	12	O56254	O56254 turnip yell
36	88	4.3	417	2	O83044	O83044 azospirillu
37	87.5	4.3	661	5	Q9NKE9	Q9nke9 drosophila
38	87.5	4.3	664	5	Q9VJW2	Q9vjw2 drosophila
39	87.5	4.3	937	2	Q9FBU8	Q9fbu8 streptomyc
40	87	4.2	620	2	Q9F3J2	Q9f3j2 streptomyc
41	87	4.2	632	5	Q9V3C9	Q9v3c9 drosophila
42	86.5	4.2	377	4	Q9H0V4	Q9h0v4 homo sapien
43	86.5	4.2	381	2	Q9WWM1	Q9wwm1 synechococc
44	86.5	4.2	445	10	Q9SZL2	Q9szl2 arabidopsis
45	86.5	4.2	577	2	Q9RYK3	Q9ryk3 deinococcus

ALIGNMENTS

RESULT	1
Q9UBY3	
ID	Q9UBY3
AC	Q9UBY3;
DT	01-MAY-2000 (Tremblrel. 13, Created)
DT	01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT	01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE	N-ACETYLGUCOSAMINE 6-O-SULFOTRANSFERASE.
GN	I-GLCNAC-6-ST.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=INTESTINE;
RX	MEDLINE=99423499; PubMed=10491328;
RA	Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;
RT	"Cloning and Characterization of a Mammalian N-Acetylglucosamine-6-
RT	Sulfotransferase that is Highly Restricted to Intestinal Tissue.";
RL	Biochem. Biophys. Res. Commun. 263:543-549(1999).
DR	EMBL; AF176839; AAD56001.1; -.
DR	EMBL; AF176838; AAD56000.1; -.
KW	Transferase.
SQ	SEQUENCE 390 AA; 43832 MW; 132D1C3B6BE38179 CRC64;

Query Match 100.0%; Score 2051; DB 4; Length 390;
Best Local Similarity 100.0%; Pred. No. 3.6e-175;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MWLPRFSSKTVTVLLLAQTTCLLLFIISRPSPSPAGGEDRVHVLVLSWSRSGSFLGQL	60
Db	1	MWLPRFSSKTVTVLLLAQTTCLLLFIISRPSPSPAGGEDRVHVLVLSWSRSGSFLGQL	60
QY	61	FSQHPDVLYLMEPAWHVWTTLSQGSAAATLHMAVRDLMKRSIFCLCDMDVFDAYMPQSRNL	120
Db	61	FSQHPDVLYLMEPAWHVWTTLSQGSAAATLHMAVRDLMKRSIFCLCDMDVFDAYMPQSRNL	120
QY	121	FFNWATSRALCSPACSAFPRGTISKQDVCKTLCIRQPFSLAREACRSYSHVVLKEVRF	180
Db	121	FFNWATSRALCSPACSAFPRGTISKQDVCKTLCIRQPFSLAREACRSYSHVVLKEVRF	180

QY	181	NLQVLYPLSDPALNLRIVHLVRDPRAVLR	SREAAGPILARDNGIVLGTNGKWVEADPHL	240
Db	181	NLQVLYPLSDPALNLRIVHLVRDPRAVLR	SREAAGPILARDNGIVLGTNGKWVEADPHL	240
QY	241	RLIREVCRSHVRIAEEAATLKPPPF	FLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL	300
Db	241	RLIREVCRSHVRIAEEAATLKPPPF	FLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL	300
QY	301	EAWIHNIHSGGIGKPIEAFHTSSRNARNV	SQAWRHALPFTKILRVQEVCGALQLLGYR	360
Db	301	EAWIHNIHSGGIGKPIEAFHTSSRNARNV	SQAWRHALPFTKILRVQEVCGALQLLGYR	360
QY	361	PVYSADQQRDLTLDLVLPRGPDHFSWASPD	390	
Db	361	PVYSADQQRDLTLDLVLPRGPDHFSWASPD	390	
RESULT	2			
Q9GZS9		PRELIMINARY;	PRT;	411 AA.
ID	Q9GZS9			
AC	Q9GZS9;			
DT	01-MAR-2001 (TReMBLrel. 16, Created)			
DT	01-MAR-2001 (TReMBLrel. 16, Last sequence update)			
DT	01-JUN-2001 (TReMBLrel. 17, Last annotation update)			
DE	INTESTINAL GLCNAC-6-SULFOTRANSFERASE (INTESTINAL N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE).			
DE	ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE).			
GN	CHST5.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20472330; PubMed=11017086;			
RA	Akama T.O., Nishida K., Nakayama J., Watanabe H., Ozaki K.,			
RA	Nakamura T., Dota A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,			
RA	Fujiwara T., Thonar E.J., Shimomura Y., Kinoshita S., Tanigami A.,			
RA	Fukuda M.N.;			
RT	"Macular corneal dystrophy type I and type II are caused by distinct			
RT	mutations in a new sulphotransferase gene.";			
RL	Nat. Genet. 26:237-241(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20472330; PubMed=11017086;			
RA	Akama T.O., Nishida K., Nakayama J., Watanabe H., Fujiwara T.,			
RA	Nakamura T., Dota A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,			
RA	Ozaki K., Kinoshita S., Shimomura Y., Tanigami A., Fukuda M.N.;			
RT	"Macular corneal dystrophy type I and type II are caused by distinct			
RT	mutations in a new sulphotransferase gene.";			
RL	Nat. Genet. 26:237-241(2000).			
DR	EMBL; AF246718; AAG28023.1; -.			
DR	EMBL; AF219991; AAG26326.1; -.			
KW	Transferase.			
SQ	SEQUENCE 411 AA; 46160 MW; 97642D54BE926E06 CRC64;			
Query Match	100.0%;	Score 2051;	DB 4;	Length 411;
Best Local Similarity	100.0%;	Pred. No. 3.9e-175;		
Matches	390;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
QY	1	MWLPRFSKTVTLLLAQTTCLLLFIISRP	GPSPAGGEDRVHVLSSWSRSGSFLGQL	60
Db	22	MWLPRFSKTVTLLLAQTTCLLLFIISRP	GPSPAGGEDRVHVLSSWSRSGSFLGQL	81
QY	61	FSQHPDVLYLMEPAWHVWTTLSQGSAA	TLHMAVRDLMRSIFLCMDVDFDAYMPQSRNLSA	120
Db	82	FSQHPDVLYLMEPAWHVWTTLSQGSAA	TLHMAVRDLMRSIFLCMDVDFDAYMPQSRNLSA	141
QY	121	FFNWATSRALCSPACSAFPRGTISKQDV	CKTLCTRPFFSLAREACRSYSHVVLEKVRFF	180
Db	142	FFNWATSRALCSPACSAFPRGTISKQDV	CKTLCTRPFFSLAREACRSYSHVVLEKVRFF	201

QY	181	NLQVLYPLSDPALNLRIVHLVRDPRAVLR	SREAAGPILARDNGIVLGTNGKWVEADPHL	240
Db	202	NLQVLYPLSDPALNLRIVHLVRDPRAVLR	SREAAGPILARDNGIVLGTNGKWVEADPHL	261
QY	241	RLIREVCRSHVRIAEEAATLKPPPF	FLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL	300
Db	262	RLIREVCRSHVRIAEEAATLKPPPF	FLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL	321
QY	301	EAWIHNIHSGGIGKPIEAFHTSSRNARNV	SQAWRHALPFTKILRVQEVCGALQLLGYR	360
Db	322	EAWIHNIHSGGIGKPIEAFHTSSRNARNV	SQAWRHALPFTKILRVQEVCGALQLLGYR	381
QY	361	PVYSADQQRDLTLDLVLPRGPDHFSWASPD	390	
Db	382	PVYSADQQRDLTLDLVLPRGPDHFSWASPD	411	
RESULT	3			
Q9GZX3		PRELIMINARY;	PRT;	395 AA.
ID	Q9GZX3			
AC	Q9GZX3;			
DT	01-MAR-2001 (TReMBLrel. 16, Created)			
DT	01-MAR-2001 (TReMBLrel. 16, Last sequence update)			
DT	01-JUN-2001 (TReMBLrel. 17, Last annotation update)			
DE	N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE GST-4BETA (CORNEAL N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE).			
DE	GST4BETA OR CHST6.			
GN	Homo sapiens (Human).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,			
RA	Rosen S.D.;			
RT	"Chromosomal Localization and Genomic Organization for the			
RT	Galactose/N-Acetylgalactosamine/N-Acetylglucosamine 6-O-			
RT	Sulfotransferase Gene Family.";			
RL	Glycobiology 0:0-0(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20472330; PubMed=11017086;			
RA	Akama T.O., Nishida K., Nakayama J., Watanabe H., Fujiwara T.,			
RA	Nakamura T., Dota A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,			
RA	Ozaki K., Kinoshita S., Shimomura Y., Tanigami A., Fukuda M.N.;			
RT	"Macular corneal dystrophy type I and type II are caused by distinct			
RT	mutations in a new sulphotransferase gene.";			
RL	Nat. Genet. 26:237-241(2000).			
DR	EMBL; AF280086; AAG48244.1; -.			
DR	EMBL; AF219990; AAG26325.1; -.			
DR	EMBL; AF219991; AAG26327.1; -.			
KW	Transferase.			
SQ	SEQUENCE 395 AA; 44098 MW; 433CA60248A48F67 CRC64;			
Query Match	84.3%;	Score 1729.5;	DB 4;	Length 395;
Best Local Similarity	85.8%;	Pred. No. 2e-146;		
Matches	333;	Conservative 17;	Mismatches 37;	Indels 1; Gaps 1;
QY	1	MWLPRFSKTVTLLLAQTTCLLLFIISRP	GPSPAGGEDRVHVLSSWSRSGSFLGQL	60
Db	1	MWLPRVSSATAVALLAQ-TFLLLFLVSR	PGSPAGGEARVHVLSSWSRSGSFFVGQL	59
QY	61	FSQHPDVLYLMEPAWHVWTTLSQGSAA	TLHMAVRDLMRSIFLCMDVDFDAYMPQSRNLSA	120
Db	60	FNQHPDVLYLMEPAWHVWTTLSQGSAA	TLHMAVRDLVRSVFLCMDVDFDAYLPWRNLSD	119
QY	121	FFNWATSRALCSPACSAFPRGTISKQDV	CKTLCTRPFFSLAREACRSYSHVVLEKVRFF	180
Db	120	LFQWAVSRALCSPACSAFPRGAISS	EAVCCKPLCARQSFLLAREACRSYSHVVLEKVRFF	179
QY	181	NLQVLYPLSDPALNLRIVHLVRDPRAVLR	SREAAGPILARDNGIVLGTNGKWVEADPHL	240

Db	180	NLQVLYPLLSDPALNLRIVHLVRDPPRAVLRSREQTAKALARDNGIVLGTNGTWVEADPGL	239
QY	241	RLIREVCRSHVRIAEAAATLKPPPLRGYRLVRFEDIAREPLAEIRALYAFTGLTLPQL	300
Db	240	RVREVCVCRSHVRIAEAAATLKPPPLRGYRLVRFEDIAREPLAEIRALYAFTGLSLTPQL	299
QY	301	EAWIHNIHGGSGIGKPIEAFTSSRNARNVSQAWRHALPFTKILRVQEVCGALQLLYR	360
Db	300	EAWIHNIHGGSGPGARAEAFKTSRNALNVSQAWRHALPFAKIRVQELCAGALQLLYR	359
QY	361	PVYSADQQRDLTLDLVLRGPDHFSWAS	388
Db	360	PVYSEDEQRNALDLVLRGLNGFTWAS	387

RESULT	4				
Q9QUP4					
ID	Q9QUP4	PRELIMINARY;	PRT;	395	AA.
AC	Q9QUP4;				
DC	01-MAY-2000	(TREMBLrel. 13, Created)			
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)			
DE	N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE.				
GN	CHST5 OR I-GLCNAC-6-ST.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6; TISSUE=INTESTINE;				
RX	MEDLINE=99423499; PubMed=10491328;				
RA	Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;				
RT	"Cloning and Characterization of a Mammalian N-Acetylglucosamine-6-				
RT	Sulfotransferase that is Highly Restricted to Intestinal Tissue.";				
RRL	Biochem. Biophys. Res. Commun. 263:543-549(1999).				
DR	EMBL; AF176841; AAD56003.1; -.				
DR	EMBL; AF176840; AAD56002.1; -.				
DR	MGD; MGI:1931825; Chst5.				
KW	Transferase.				
SQ	SEQUENCE 395 AA; 44537 MW; 3FDF71E43ED383BE CRC64;				

[illegible]

```

Db      360 SVHSELEQRDLSLDLLPRGMDSEKNAS 387

RESULT      5
Q9Y5R3
ID  Q9Y5R3      PRELIMINARY;      PRT;      386 AA.
AC  Q9Y5R3;
DI  01-NOV-1999 (TREMBLrel. 12, Created)
DT  01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT  01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE  N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE (L-SELECTIN LIGAND
DE  SULFOTRANSFERASE GST-3).
GN  GST3.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=TONSIL;
RX  MEDLINE=99264336; PubMed=10330415;
RA  Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,
RA  Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.;
RT  "Sulfotransferases of two specificities function in the reconstitution
RT  of high endothelial cell ligands for L-selectin.";
RL  J. Cell Biol. 145:899-910(1999).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  TISSUE=TONSIL;
RA  Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,
RA  Rosen S.D.;
RT  "Chromosomal Localization and Genomic Organization for the
RT  Galactose/N-Acetylgalactosamine/N-Acetylglucosamine 6-O-
RT  Sulfotransferase Gene Family.";
RL  Glycobiology 0:0-0(2001).
DR  EMBL; AF131235; AAD33015.1; -.
DR  EMBL; AF280088; AAG48246.1; -.
DR  InterPro; IPR003015; HLH_MYC.
DR  PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
KW  transferase; Lectin; Selectin.
SQ  SEQUENCE      386 AA; 45133 MW;  0C3BB4022417143A CRC64;

```

```

Query Match      49.1%; Score 1008; DB 4; Length 386;
Best Local Similarity 52.3%; Pred. No. 6e-82;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

QY 14 LLLAQTCLLLFIISRP-----GPSSPAGGEDRVHVLVLSSWRSGSSELGQ 59
Db 1 MLLPKKMKLLLFVQSMAILLALPFHMYSHNISSLKMAQPERMHVLVLSSWRSGSSEVQG 60

QY 60 LFSQHPDVLYLMEPAWHVWVTTLSQGSAAATLHMAVRDLMRSLFLCDMDVFDAYM-PQSRNL 118
Db 61 LFGQHPDVLYLMEPAWHVWMTFKQSTAWMLHMAVRDLTRAVFLCDMSVFDAYMEPGPRQ 120

QY 119 SAFFNWATSRALCSPACSAFPRGTISKQDVCKTLCTRQPFSLAREACRSYSHVVLKEVR 178
Db 121 SSLFQWENSRALCSAPACDIIPQDEIIPRAHCRLLCSQQPFVEVEKACRSYSHVVLKEVR 180

QY 179 FFNQLVLYPLLSDFALNLRVHLVRDPRAVLRSREAAAGPILARDNGIVLGTN-GKWVEAD 237
Db 181 FFNQLVLYPLLKDPSLNLHVHLVRDPRAVFRSRERTKGDLMIDSRIVMGQHEQKLKED 240

QY 238 PHRLIREVCRSHVRIAEAAATLKP-PPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTL 296
Db 241 QPYVVMQVICSQLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRMFYFVGLF 298

QY 297 TPQLEAWIHNITHSGGIGKPIEAFHTSSRNAFNVSQAWRHALPFTKILRVQEVCGALQL 356
Db 299 LPHLQTVHVNITRGKMGMD--HAFHTNARDALNVSQAWRWSLPYEKVSRLQKACGDAMNL 356

QY 357 LGRPVYSADQQORDLTLDLVLPRGPDHFSWASPD 390
Db 357 LGRPVYSADQQORDLTLDLVLPRGPDHFSWASPD 390

```

Db	357	LGYRHRSEQEQRNLLLDLL-----STWTVPE	383
RESULT	6		
Q9R1I1			
ID	Q9R1I1	PRELIMINARY;	PRT; 388 AA.
AC	Q9R1I1;		
DT	01-MAY-2000 (TrEMBLrel. 13, Created)		
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)		
DE	L-SELECTIN LIGAND SULFOTRANSFERASE.		
GN	CHST4.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99361934; PubMed=10435581;		
RA	Hiraoka N., Petryniak B., Nakayama J., Tsuboi S., Suzuki M., Yeh J.C.,		
RA	Izawa D., Tanaka T., Miyasaka M., Lowe J.B., Fukuda M.;		
RT	"A novel, high endothelial venule-specific sulfotransferase expresses		
RT	6-sulfo sialyl Lewis X, an L-selecitin ligand displayed by CD34.";		
RL	Immunity 11:79-89(1999).		
DR	EMBL; AF109155; AAD45579.1; -.		
DR	MGD; MGI:1349479; Chst4.		
KW	Transferase; Lectin; Selectin.		
SQ	SEQUENCE 388 AA; 44635 MW; 6D5371AFB6884AEE CRC64;		
Query Match	47.1%;	Score 965.5;	DB 11; Length 388;
Best Local Similarity	55.9%;	Pred. No. 3.8e-78;	
Matches	194;	Conservative 41;	Mismatches 87; Indels 25; Gaps 6;
QY	42	VHVLVSSWRSGSFLGQLFSQHPDVVFYLMPEAWHVWTTLSQGSAAATHMAVRDLMSIF	101
Db	42	VHVLVSSWRSGSFLGQLFSQHPDVVFYLMPEAWHVWTTLSQGSAAATHMAVRDLMSIF	101
QY	102	LCMDVFDAYM-PQSRNLSAFFNWATSRALCSPACSAFPRGTISKQDVCKTLCTRPQFS	160
Db	102	LCDSMVFDAYMNPGRKQSSLFQWESRALCSAPVCDFEPPAHEISSPKHCKLLCGQPFED	161
QY	161	LAREACRSYSHVVLKEVREFNLQVLYPLSDPALNLRIVHLVRDPRAVLRSREAAGPILA	220
Db	162	MVEKACRSHGFVVLKEVRFSLQALYPLTDPNLHVHLVRDPRAVRSREHTTIELV	221
QY	221	RDNGIVLGTNGKWEADPHLRLE-----VCRSHVRIAEE-ATLKPPPFELRGY	269
Db	222	VDSHIVLG-----QHLETIKEEDOPYYAMKIICKSQVDIVKAIQTL--PEALQORY	270
QY	270	RLVRFEDLAREPLAEIRALYAFTGLTLPQLEAWIHNITHGSGIGKPIEAFHTSSRNARN	329
Db	271	LFLRYEDLVRAPLAQTTRLYKFVGLDFPLHLQTVWVHNVRGKGMGQ--HAFHTNARNALN	328
QY	330	VSOAWRHALPFTKILRVQEVCAQALQLLGYRPPVYSADQQRDLTLDLV	376
Db	329	VSOAWRWSLPEYKVSQIQDACGEAMDLLGYLQVRSQQEQGNLSLDLL	375
RESULT	7		
Q9WUE5			
ID	Q9WUE5	PRELIMINARY;	PRT; 388 AA.
AC	Q9WUE5;		
DT	01-NOV-1999 (TrEMBLrel. 12, Created)		
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)		
DE	N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE (CARBOHYDRATE (CHONDROITIN		
DE	6/KERATAN) SULFOTRANSFERASE 4).		
GN	CHST4.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
[1]			
RN	SEQUENCE FROM N.A.		
RP	STRAIN=C57BL/6J; TISSUE=TONGUE;		
RX	MEDLINE=21085660; PubMed=11217851;		
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,		
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,		
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,		
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,		
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,		
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,		
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,		
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,		
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,		
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,		
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,		
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,		
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,		
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,		
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,		
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,		
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,		
RA	Hayashizaki Y.;		
RT	"Functional annotation of a full-length mouse cDNA collection.";		
RL	Nature 409:685-690(2001).		
DR	EMBL; AF131236; AAD33016.1; -.		
DR	EMBL; AK009113; BAB26078.1; -.		
DR	MGD; MGI:1349479; Chst4.		
KW	Transferase.		
SQ	SEQUENCE 388 AA; 44694 MW; D1E9D7796DF8574D CRC64;		
Query Match	46.7%;	Score 958.5;	DB 11; Length 388;
Best Local Similarity	55.6%;	Pred. No. 1.6e-77;	
Matches	193;	Conservative 42;	Mismatches 87; Indels 25; Gaps 6;
QY	42	VHVLVSSWRSGSFLGQLFSQHPDVVFYLMPEAWHVWTTLSQGSAAATHMAVRDLMSIF	101
Db	42	VHVLVSSWRSGSFLGQLFSQHPDVVFYLMPEAWHVWTTLSQGSAAATHMAVRDLMSIF	101
QY	102	LCMDVFDAYM-PQSRNLSAFFNWATSRALCSPACSAFPRGTISKQDVCKTLCTRPQFS	160
Db	102	LCDSMVFDAYMNPGRKQSSLFQWESRALCSAPVCDFEPPAHEISSPKHCKLLCGQPFED	161
QY	161	LAREACRSYSHVVLKEVREFNLQVLYPLSDPALNLRIVHLVRDPRAVLRSREAAGPILA	220
Db	162	MVEKACRSHGFVVLKEVRFSLQALYPLTDPNLHVHLVRDPRAVRSREHTTIELM	221
QY	221	RDNGIVLGTNGKWEADPHLRLE-----VCRSHVRIAEE-ATLKPPPFELRGY	269
Db	222	VDSHIVLG-----QHLETIKEEDOPYYAMKIICKSQVDIVKAIQTL--PEALQORY	270
QY	270	RLVRFEDLAREPLAEIRALYAFTGLTLPQLEAWIHNITHGSGIGKPIEAFHTSSRNARN	329
Db	271	LFLRYEDLVRAPLAQTTRLYKFVGLDFPLHLQTVWVHNVRGKGMGQ--HAFHTNARNALN	328
QY	330	VSOAWRHALPFTKILRVQEVCAQALQLLGYRPPVYSADQQRDLTLDLV	376
Db	329	VSOAWRWSLPEYKVSQIQDACGEAMDLLGYLQVRSQQEQGNLSLDLL	375
RESULT	8		
O88276			
ID	O88276	PRELIMINARY;	PRT; 530 AA.
AC	O88276;		

Query Match		32.2%;	Score 659.5;	DB 4;	Length 530;	
Best Local Similarity		39.1%;	Pred. No. 1.3e-50;			
Matches 151;		Conservative	68;	Mismatches 122;	Indels 45;	Gaps 12;
QY	31	GPSSPAG-----	GEDRVHVLVLSWRS	SGSFLGQLFSQHDPDVFYLM	EPAWHVWTTL	81
		::	::	::	::	::
Db	145	GVAAPPNGTRGTGGV	GDKRLVYVFTTWRS	GSFFGELFNQNP	EVFFLYEPVHVWQKL	204
		::	::	::	::	::
QY	82	SGGSAATLHMAVRDL	MRSIFLCDMDVFDAYMPQ---	SRNLS--	AFFNWATSRALCSP	136
		::	::	::	::	::
Db	205	YPGDAVSLQGAARD	MLSALYRCDLSVFQLYSPAGSGGRNLT	TTLGIFGAATNKVVCSSPLC	264	
		::	::	::	::	::
QY	137	SAFPRGTISKQD--	VCKTILCTROPFSLAREACRSYSHVVLKEV	REFFNLQVLYPL	SDPAL	194
		::	::	::	::	::
Db	265	PAYRKEVVG	LVDDRCKK-CPPQRLARFEE	CKRYRTLVIKGRV	FDVAVLAPLLRDPAL	323
		::	::	::	::	::
QY	195	NLRIVHLVRDP	RAVLRS-----	REAAGPILARD-----	NGIVLGTNGKWV	234
		::	::	::	::	::
Db	324	DLKVIHLVRDP	RAVASSRIRSRHGLIRES	LQVRSRDPRAHRMPF	LEAAGHKLGAKEGV	383
		::	::	::	::	::
QY	235	--EADPH-LRLIRE	VCRRSHVRIAEAAATLKPP	FLRGYRLVR	FEDLAREPLAEIRALYAF	291
		::	::	::	::	::
Db	384	GGPADYHALGAME	VICNSMAKTLQTA-LOPPD	WLQGHYLVRYED	LVGDPVKTLRRVDF	442
		::	::	::	::	::
QY	292	TGLTLTPQLEAW	IHNITHGSG-IGKPIEAFHTSSRNARNV	SQAWRHALPFTKILRVQ	EVVC	350
		::	::	::	::	::
Db	443	VGLLVSPEME	QFALNMTSGGSSSKP--	FVVSARNATQAA	NAWRTALTITFQOIKQVEEFC	499
		::	::	::	::	::
QY	351	AGALQLLGYRP	VPYSADQQORDLTLDLV	376		
		::	::	::	::	::
Db	500	YQPM	AVLGYERN	SPPEVKDLSKTL	525	
		::	::	::	::	::
RESULT 11						
Q9UED5						
ID	Q9UED5	PRELIMINARY;		PRT;	484	AA.
AC	Q9UED5;					
DT	01-MAY-2000	(TReMBLrel. 13, Created)				
DT	01-MAY-2000	(TReMBLrel. 13, Last sequence update)				
DT	01-MAY-2000	(TReMBLrel. 13, Last annotation update)				
DE	N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE (GLCNAC6ST).					
GN	GN6ST.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
OX	NCBI_TaxID=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	TISSUE=BRAIN;					
RA	Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,					
RA	Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,					
RA	Ozaki T., Nakagawara A., Kadomatsu K., Muramatsu T.;					
RT	"Human N-acetylglucosamine-6-O-sulfotransferase involved in the					
RT	biosynthesis of 6-sulfo sialyl Lewis X: Molecular cloning, chromosomal					
RT	mapping, and expression in various organs and tumor cells.";					
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.					
DR	EMBL; AB014679; BAA34265.1; -.					
KW	Transferase.					
SQ	SEQUENCE 484 AA; 53641 MW; D077EB	CD7645F2F0 CRC64;				
Query Match		32.0%;	Score 656.5;	DB 4;	Length 484;	
Best Local Similarity		39.7%;	Pred. No. 2.1e-50;			
Matches 149;		Conservative	68;	Mismatches 121;	Indels 37;	Gaps 12;
QY	34	SPAGGEDRVH-VLV	SSWRS	SGSFLGQLFSQHDPDVFYLM	EPAWHVWTTLSQGS	AATLHMA 92
		::	::	::	::	::
Db	110	APEGVGDKRHWMYVFTTWRS	GSFFGELFNQNP	EVFFLYEPVHVWQKL	YPGD	AVSLQGA 169
		::	::	::	::	::
QY	93	VRDLMRSTFLCDMDVFDAYMPQ---	SRNLS--	AFFNWATSRALCSP	PACSAFPRGTISKQ	147
		::	::	::	::	::
Db	170	ARDMLSALYRCDLSVFQLYSPAGSGGRNLT	TTLGIFGAATNKVVCSSPLC	PAYRKEVVG	LV	229
		::	::	::	::	::

QY	148	D--VCKTILCTROPFSLAREACRSYSHVVLKEV	REFFNLQVLYPL	SDPALNLRIVHLVRDP	205	
Db	230	DDRCKK-CPPQRLARFEECRKYRTLVIKGRV	FDVAVLAPLLRDPALDKVIHLVRDP	288		
QY	206	RAVLRS-----	REAAGPILARD-----	NGIVLGTNGKWV--	EADPH-LRL 242	
Db	289	RAVASSRIRSRHGLIRESLQVRSRDPRAHRMP	PFLEAAGHKLGAKEGVGGPADIHALGA	348		
QY	243	IREVCRSHVRIAEAAATLKPPPFRLGRYRLVR	FEDLAREPLAEIRALYAF	TGLTLTPQLEA	302	
Db	349	MEVICNSMAKTLQTA-LQPPDWLQGHYLVVRY	EDLVGDPVKTLRRYVDFVGLLVSP	EMEQ	407	
QY	303	WIHNITHGSG-IGKPIEAFHTSSRNARNV	SQAWRHALPFTKILRVQEV	CAGALQLLGYRP	361	
Db	408	FALNMTSGSGSSKP--	FVVSARNATQAA	NAWRTALTITFQOIKQVEEFCYQPM	AVLGYER 464	
QY	362	VYSADQQORDLTLDLV	376			
Db	465	VNSPEEVKDL	SKTLL 479			
RESULT 12						
QY	Q9Y4C5	PRELIMINARY; PRI; 531 AA.				
AC	Q9Y4C5;					
DT	01-NOV-1999	(TReMBLrel. 12, Created)				
DT	01-NOV-1999	(TReMBLrel. 12, Last sequence update)				
DT	01-NOV-1999	(TReMBLrel. 12, Last annotation update)				
DE	LONG FORM OF N-ACETYGLUCOSAMINE-6-O-SULFOTRANSFERASE (GLCNAC6ST).					
GN	GN6ST.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
OX	NCBI_TaxID=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	TISSUE=BRAIN;					
RA	Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,					
RA	Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,					
RA	Ozaki T., Nakagawara A., Kadomatsu K., Muramatsu T.;					
RT	"Human N-acetylglucosamine-6-O-sulfotransferase involved in the					
RT	biosynthesis of 6-sulfo sialyl Lewis X: Molecular cloning, chromosomal					
RT	mapping, and expression in various organs and tumor cells.";					
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.					
DR	EMBL; AB014680; BAA34266.1; -.					
KW	Transferase.					
SQ	SEQUENCE	531 AA;	58768 MW;	D2EDB74E95B5162F	CRC64;	
Query Match 32.0%; Score 656.5; DB 4; Length 531;						
Best Local Similarity 39.7%; Pred. No. 2.4e-50;						
Matches 149; Conservative 68; Mismatches 121; Indels 37; Gaps 12;						
QY	34	SPAGGEDRVH-VLV	SSWRS	SGSFLGQLFSQHDPDVFYLM	EPAWHVWTTLSQGS	AATLHMA 92
Db	157	APEGVGDKRHWMYVFTTWRS	GSFFGELFNQNP	EVFFLYEPVHVWQKL	YPGD	AVSLQGA 216
QY	93	VRDLMRSIFLCDMDVFDAYMPQ----	SRNLS--	AFFNWATSRALCSP	PACSAFPRGTISKQ	147
Db	217	ARDMLSALYRCDLSVFQLYSPAGSGGRNLT	TTLGIFGAATNKVVCSSPLC	PAYRKEVVG	LV	276
QY	148	D--VCKTILCTROPFSLAREACRSYSHVVLKEV	REFFNLQVLYPL	SDPALNLRIVHLVRDP	205	
Db	277	DDRCKK-CPPQRLARFEECRKYRTLVIKGRV	FDVAVLAPLLRDPALDKVIHLVRDP	335		
QY	206	RAVLRS-----	REAAGPILARD-----	NGIVLGTNGKWV--	EADPH-LRL 242	
Db	336	RAVASSRIRSRHGLIRESLQVRSRDPRAHRMP	PFLEAAGHKLGAKEGVGGPADIHALGA	395		
QY	243	IREVCRSHVRIAEAAATLKPPPFRLGRYRLVR	FEDLAREPLAEIRALYAF	TGLTLTPQLEA	302	
Db	396	MEVICNSMAKTLQTA-LQPPDWLQGHYLVVRY	EDLVGDPVKTLRRYVDFVGLLVSP	EMEQ	454	

Db

263

RDPEGLNKVVQLFRDPRAVHNSRLKSRQGLLRESIQVLRTRQRGDFHRVLLAHGVDDARP

322

|| |||:| | ||||| || : | |: || | : | :| |

QY

239

--HLRLREVCRRSHVRIAEAA-----TLKPPPFRLGRYRLVRFEDLAREPLA

283

| : | : | : | : | : | : | : | : | : | : | : | : |

Db

323

GGQARALPSAPRADFFLTSALEVICEAWLRLDLLFTRGAPAWLRRRLRYEDLVWQPQA

382

| : | : | : | : | : | : | : | : | : | : | : | : |

QY

284

EIRALYAFTGLTLPQLEAWIHNIHNSGSGICKPIEAFHTSSRNARNVSOAWRHALPFTKI

343

: : | | | : | : | : | : | : | : | : | : | : | : |

Db

383

QLRRLRLRFSGLRFTLAALDAFAFNMTGRSGAYGAD-RPFHLSARDAREAVHAWRERLSQEQV

441

: : | | | : | : | : | : | : | : | : | : | : | : |

QY

344

LRVQEVCCAGALQLLGY

359

: : | | | : | : | : | : |

Db

442

QVETACAPAMRLLAY

457

: : | | | : | : | : | : |

QY

341

TKILRVQEVCCAGALQLLGY

359

: : | : | : | : | : | : | : |

Db

441

EQVRQVEAACAPAMRLLAY

459

: : | : | : | : | : | : | : |

Search completed: January 31, 2002, 09:50:43
Job time: 331 sec

RESULT 15

075667

PRELIMINARY;

PRT;

486 AA.

AC

075667;

01-NOV-1998 (TremBLrel. 08, Created)

DT

01-NOV-1998 (TremBLrel. 08, Last sequence update)

DT

01-MAR-2001 (TremBLrel. 16, Last annotation update)

DE

DJ71L16.4 (N-ACETYLGUCOSAMINE-6-O-SULFOTRANSFERASE).

GN

DJ71L16.4.

OS

Homo sapiens (Human).

OC

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX

NCBI_TaxID=9606;

RN

[1]

RP

SEQUENCE FROM N.A.

RA

Grafham D.;

RL

Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

RN

[2]

RP

SEQUENCE FROM N.A.

RC

TISSUE=BRAIN;

RA

Uchimura K., Muramatsu T.;

RT

"Identification and molecular characterization of a cDNA encoding a

RT

novel N-acetylglucosamine-6-O-sulfotransferase.";

RL

Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR

EMBL; AL022165; CAA18154.1; -.

DR

EMBL; AB040711; BAB13770.1; -.

KW

Transferase.

SQ

SEQUENCE 486 AA; 54266 MW; 3F1FD1430B3C8E95 CRC64;

Query Match

29.7%;

Score 610;

DB 4;

Length 486;

Best Local Similarity

38.8%;

Pred. No. 3e-46;

Matches 147;

Conservative 52;

Mismatches 126;

Indels 54;

Gaps 11;

QY

30

PGPSSPAGGE---DRHVVLVLSWSRSGSFLGQLFSQHPDVLYLMEPAWHVWTLTSLQGS

85

| | | | | : : | : | : | : | : | : | : | : | : | : | : |

Db

86

PSNLSGAVGEAVSREKQHIYVHATWRTGSSFLGELFNQHPDVLYLMEPMWHLWQALYPGD

145

| | | | | : : | : | : | : | : | : | : | : | : | : | : |

QY

86

AATLHMAVRDLMRSLFCLDMDFDAY-----MPQSRNL--SAFFNWATSRALCSPP

134

| : | : | : | : | : | : | : | : | : | : | : | : | : |

Db

146

AESLQGALRDLRLSLFRCDPSVRLYAPPGDPAARAPDTANLTAALEFRWRTNKVICSP

205

| : | : | : | : | : | : | : | : | : | : | : | : | : |

QY

135

ACSAPFRGTISKQDVCKTLCTRQ--PFSL-AREA-CRSYSHVVLKEVRFNQLVLYPLLS

190

| | | | | | | | | | | | | | | | | | | | | | | | |

Db

206

LCPGAPRARAEEVGLVEDTACERSCPPVAIRALEAECRKYPVVKDVRLDLGLVLPCLR

265

| | | | | | | | | | | | | | | | | | | | | | | | |

QY

191

DPALNLRIVHLVRDPA-----VLSREAAG-----PILARDNGIVLG

228

| | | | | : | : | : | : | : | : | : | : | : | : |

Db

266

DPGLNLKVVQLFRDPRAVHNSRLKSRQGLLRESIQVLRTRQRGDFHRVLLAHGVGARPG

325

| | | | | : | : | : | : | : | : | : | : | : | : |

QY

229

TNGKWVEADPH-----LRLIREVCRRSHVRIAEAAATLKPPPLRGRYRLVRFEDLARE

280

: : | | | | | : | : | : | : | : | : | : | : | : |

Db

326

GQSRALPAAPRADFFLTGALEVICEAWLRLDLLFARGA---PAWLRRYRLRYEDLVQR

381

: : | | | | | : | : | : | : | : | : | : | : | : |

QY

281

PLAEIRALYAFTGLTLPQLEAWIHNIHNSGSGICKPIEAFHTSSRNARNVSOAWRHALPF

340

| : : | | | : | : | : | : | : | : | : | : | : | : |

Db

382

PRAQLRRLRLRFSGLRALAALDAFALNMTRGAAYGAD-RPFHLSARDAREAVHAWRERLSR

440

| : : | | | : | : | : | : | : | : | : | : | : | : |


```

RESULT 2
JC7350
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - mouse
C;Species: Mus musculus (house mouse)
C;Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000
C;Accession: JC7350
R;Uchimura, K.; Fasakhany, F.; Kadamatsu, K.; Matsukawa, T.; Yamakawa, T.; Kurosawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A;Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
A;Reference number: JC7350; MUID:20374462
A;Accession: JC7350
A;Molecule type: mRNA
A;Residues: 1-484 <UCH>
A;Cross-references: DDBJ:AB040710
C;Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
C;Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 29.9%; Score 613.5; DB 2; Length 484;
Best Local Similarity 38.8%; Pred. NO. 2.1e-47;
Matches 146; Conservative 54; Mismatches 127; Indels 49; Gaps 11;

QY 30 PGPSSPAG--GEDRVHVLVLSSWRSGSFLGQLFSDHPDVFYLMPEAWHVWTLTSLQSGSA 86
DDB 85 PGNLSAVGEAVTQEKQHIYVHATWRTGSSFLGELFNPDPVFYLYEPMWHLWQALYPGDA 144
QY 87 ATLHMAVRDLMRISFLCDMDVFDAY-----MPOSRLSA--FFNWAISRALCSPPA 135
DDB 145 ESLOGLARDMLRSLFRCDPSVRLYAQPGDGERAPDSANLTAMLFWRRTKNKVICSPPL 204
QY 136 CSAFRGT---ISKQDVCKTICTROPFSL-AREA-CRSYSHVVLKEVRFNQLQVLPPL 189
DDB 205 CPAAPRARADVGLVEDKACESTC--PPVSLRALEAECRKYPVVVKDVRLLDLGLVLPPL 262
QY 190 SDPALNLRIVHLVRDPRAVLRSREAAAGPILARDNGIVLGNGK-----WVEADP 238
DDB 263 RDPGLNLKVVQLFRDPRAVHNSRLKSRQGLLRESIQVLRTRQRGDHFRVLLAHGVDPAR 322
QY 239 --HLRLIREVCRSHVRIAEAA-----TLKPPPELRGRYRLVRFEDLAREPLA 283
DDB 323 GGQARALPSAPRADFFLTSALEVICEAWLRDLFTRGAPAWLRRYRLRLRYEDLVWQPOA 382
QY 284 EIRALYAFGLTLTPQLEAWIHNIHSGIGKPIEAFHTSSRNARNVSVQAWRHLPFTKI 343
DDB 383 QLRRLLRFSGLTLAALDAFAFNMTGRSAYGAD-RPFHLSARDAREAVHVRERLSQEQV 441
QY 344 LRQVEVCAGALQLLGY 359
DDB 442 RQVETACAPAMRLLAY 457

RESULT 3
JC7351
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - human
C;Species: Homo sapiens (man)
C;Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000
C;Accession: JC7351
R;Uchimura, K.; Fasakhany, F.; Kadamatsu, K.; Matsukawa, T.; Yamakawa, T.; Kurosawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A;Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
A;Reference number: JC7350; MUID:20374462
A;Accession: JC7351
A;Molecule type: mRNA
A;Residues: 1-486 <UCH>
A;Cross-references: DDBJ:AB040711
C;Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
C;Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

```

Db 336 RQPRWLRGRYMLVRYEDVARAPLRKALEMYRFAGIHTPQVEEWIRANTQAP---QDSNG 392
QY 320 FHTSSRNARNVSOAWRHALPFTKILRVQEVACALQLLGYRPVYSADQQRDLDLVLPR 379
Db 393 IYSTQKNSSEQFEKWRFSIPFKLAQVVQDACEPAMRLFGYKLAASSAQELTNRSLSL-LEE 451
QY 380 GP 381
Db 452 GP 453

RESULT 5
TI6350
hypothetical protein F42G9.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jan-2000
C:Accession: T16350
R;Taich, A.
submitted to the EMBL Data Library, March 1996
A:Description: The sequence of C. elegans cosmid F42G9.
A:Reference number: Z18498
A:Accession: T16350
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-359 <TAI>
A:Cross-references: EMBL:U00051; NID:g1216305; PID:g1216308; PIDN:AAA91354.1; CESP:F42G9
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:F42G9.8
A:Introns: 28/2; 50/2; 113/2; 155/3; 220/3; 290/2
C:Superfamily: Caenorhabditis elegans hypothetical protein F42G9.8

Query Match 5.1%; Score 105; DB 2; Length 359;
Best Local Similarity 19.8%; Pred. No. 0.11;
Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;
QY 11 VTVLLAQTTCLLFI-----ISRPGSPSPAGGEDRVH-----VLV 46
Db 31 IYIFIFCTICLLIFSSIKCKKLEQKLEQLSLSKESLIFNQDARHSRLLSNLEQLIFV 90
QY 47 LSSWRSGSFLGQFSQHPDV-----FYLMEPAW-HVWTTLSQGSAAATLHMAVR 94
Db 91 GGVPRSGTTLMRAILDAPHDPVRCGGETMLLPSTFTWQAGWRNDWV---NNSGIT----- 141
QY 95 DLMRSIFLCMDVFDAYMPQSRNLSAFFNWATSRALCSPACSAFPRGIISK-QDVCKTL 153
Db 142 -----QEVFD-----DAVSAFIIEIVAKHSELAPRL 167
QY 154 CTRQPFSLAREACRSYSHVVLKEVRFNQLVPLSDPALNLRIVHLVYRDPRAVLRSR- 212
Db 168 CNKDP-----YTALWLPTR-----RLYP-----NAKFILMIRDAVAVHSMI 205
QY 213 EAAGPILARDNGIVLTNGKWEADPHRLRIREVCRSHVRIAEAAATLKPPPLFGRYRLV 272
Db 206 ERKVPVAGYNTSDEISMVQW---NOELRKMTFFQCNN---APQCCK-----V 247
QY 273 RFEDLAREPLAEIRALYAFTGLTLTPQLEAWIHNITHGSGIGKPI-----EAFHTSS-RNA 327
Db 248 YIERLIQKPAEELIRITNFDLPFSQOM-----LRHQDLIGDEVLDNQEFSSASQVKNS 301
QY 328 RNVS--QAWRHALPFTKILRVQEVACALQLLGY-----RPVYS 364
Db 302 INTKALTSWEDCFSEETLRKLDV-AFPLGLGYDTSISKPDYS 344

RESULT 6
G70729
hypothetical protein Rv2267c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70729
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: G70729
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-388 <COL>
A:Cross-references: GB:Z77163; GB:AL123456; NID:g3261610; PIDN:CAB00968.1; PID:e25507
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv2267c

Query Match 5.0%; Score 102.5; DB 2; Length 388;
Best Local Similarity 22.7%; Pred. No. 0.2;
Matches 66; Conservative 37; Mismatches 99; Indels 89; Gaps 15;
QY 44 VLVLSWRSGSFLGQLF---SQH-----PDVFYLMPEAHVWVTTLSQGSAAATLH 90
Db 83 IFIVGHWRTGTLLHELLVVDRTGPTGYECLAPHHLLTE-----WFAPYVEFLVSKH 137
QY 91 MAVRDLMRISIFLCMDVFDAYMPQSRNLSAFFNWATSRALCSPACSAFPRGTISKQDVC 150
Db 138 RAMDNMDLSL-----HHPOEDE----FWC-MQGLPSPYLTIAPP----- 172
QY 151 KTLCTRQP-----FSLAREACRSYSHVVLKEVRFNQLVLY-----PLLSDPALNLR- 198
Db 173 ---NRPPQYEEYLDLEQVAPRELE--IWKRTLFRFVQVYFRRRKTVILKNPTHSFRIK 226
QY 199 -----VHLVRDPRAVLRSRREAAAGPILARDNGIVL---GTNGKWEADPHRLRI 243
Db 227 VLLEVPQAKFIHIVRDPYVVPSTIHLHKALYRIHGLQQPTFDGLDDKVVST--YVDLY 284
QY 244 REVCRSHVRIAEAAATLKPPPLFGRYRLVRFEDLAREPLAEIRALYAFTGL 294
Db 285 R-----KLDEGRELVDPDPT---RFVELRYEDLIGDPEGQLRRLYQHILGL 324

RESULT 7
DEHUPT
pyruvate dehydrogenase (lipoamide) (EC 1.2.4.1) alpha chain precursor, testis-specific
N/Alternate names: pyruvate dehydrogenase complex, E1 component alpha chain
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 15-Oct-1999
C:Accession: A37104
R;Dahl, H.H.M.; Brown, R.M.; Hutchison, W.M.; Maragos, C.; Brown, G.K. Genomics 8, 225-232, 1990
A:Title: A testis-specific form of the human pyruvate dehydrogenase Elalpha subunit i
A:Reference number: A37104; MUID:91065637
A:Accession: A37104
A:Molecule type: mRNA
A:Residues: 1-388 <DAH>
A:Cross-references: GB:M86808; GB:J04769; NID:g190789; PIDN:AAA60232.1; PID:g190790
C:Genetics:
A:Gene: GDB:PDHA2
A:Cross-references: GDB:I20711; OMIM:179061
A:Map position: 4q22-4q23
C:Superfamily: pyruvate dehydrogenase (lipoamide) alpha chain; thiamin pyrophosphate-
C:Keywords: flavoprotein; heterotetramer; mitochondrion; oxidoreductase; phosphoprote
F;1-27/Domain: transit peptide (mitochondrion) #status predicted <TNP>
F;28-388/Product: pyruvate dehydrogenase (lipoamide) alpha chain #status predicted <M
F;183-232/Domain: thiamin pyrophosphate-binding domain homology <TPB>
F;230/Binding site: phosphate (Ser) (covalent) #status predicted
F;291/Binding site: phosphate (Ser) (covalent) #status predicted
F;298/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 4.7%; Score 96; DB 1; Length 388;
Best Local Similarity 23.2%; Pred. No. 0.76;
Matches 45; Conservative 25; Mismatches 82; Indels 42; Gaps 7;

QY	82	SQGS--AATLHMAVRDLMR	SIFLCMDVFDAYMPQSRNLSA	FFNWATSRALCSP	PACSAFP	140
		:		:	:	:
Db	198	NQQIAEAFNMAALWKLPCV	FICENNLYG-----MGTSTERA	AAASP----	DYK	242
		:	:	:	:	:
QY	141	RGTI---SKQDVCKTLCT	ROPFSLAREACRSYSHVV	LKEVRFFNLQVLYPL	LLSDPALNLR	197
		:		:	:	:
Db	243	RCNFIPGLKVDGMDVLCV	REATKFAANYCRSGKGP	ILMELQTYRYH--GH	SMSPGVSYR	300
		:		:	:	:
QY	198	I-----VHLVRDPRAVL	RSREAGPILARDNGIV	LGTNGKWVEADPHLRL	IREVCRSHV	251
			:	:	:	:
Db	301	TREIQEVRSKRDP	IIILQDRM-----VNSK	LATVEELKEIGA	EVURKEID	345
		:	:	:	:	:
QY	252	RIAEAAATLKP	PPFL	265		
		:				
Db	346	DAACFATTD	PEPHL	359		
		:				

RESULT 8
S29029 phenylalanine ammonia-lyase (EC 4.3.1.5) - sweet potato
C:Species: Ipomoea batatas (sweet potato)
C:Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C:Accession: S29029
R:Tanaka, Y.; Matsuoka, M.; Yamanoto, N.; Ohashi, Y.; Kano-Murakami, Y.; Ozeki, Y.
Plant Physiol. 90, 1403-1407, 1989
A:Title: Structure and characterization of a cDNA clone for phenylalanine ammonia-lyase
A:Reference number: S29029
A:Accession: S29029
A:Molecule type: mRNA
A:Residues: 1-707 <TAN>
A:Cross-references: EMBL:N29232; NID:g168271; PIDN:AAA33389.1; PID:g168272
C:Superfamily: histidine ammonia-lyase
C:Keywords: ammonia-lyase; carbon-nitrogen lyase
F:192-194/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F:193/Modified site: dehydroalanine (Ser) #status predicted

Query Match	4.7%;	Score 96;	DB 2;	Length 707;	
Best Local Similarity	25.1%;	Pred. No. 1.7;			
Matches 57;	Conservative 26;	Mismatches 86;	Indels 58;	Gaps 10;	

QY	159	FSLAREACRSYSHVVLKE-----VREFNLQVLYPLLS---DPALNLRIVH	200
		: : : : : : : : : : :	
Db	131	FGNATESCHTLPHSATRAAMLVRINTLLQYSGIRFEILEATKLLNHNITPCLPLR--G	188
QY	201	LVRDPRAVLRSRRAACPIIARDNGIVLGTNGKWEADPHRLRIREVCRSHVRIAEAA---	257
		: : : : : : : : : : : : :	
Db	189	TITASGDLVPLSYIAGLTITGRPNASKAVGPNGETLNAEEALKL---SRSGRRIFRVASPR	244
QY	258	---TLKPPPFRLGRYRLVRFEDLAREPLAEI-RALYA-----FTGLTLTPQLE---	301
		: : : : : : : : : : : : : :	
Db	245	KGLPSLMAPPVLVGLMASMVLFEEANVLAVLSEVLSAIFAEVNMGKPEFTD-HLTHKLIKHP	303
QY	302	-----AWIHNITHSGIGKPIEAFH-----TSSRNARNVSOAW	334
		: : : : : : : : : : : :	
Db	304	GOTEFAAAITMEHILDGSSYVKAACKLHEMDPLOKPKQDRYALRTSPQW	350

RESULT 9
T38905
probable valine-pyruvate transaminase (EC 2.6.1.66) SPAC56E4.03 [similarity] - fission
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000
C/Accession: T38905
R/Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A/Reference number: Z21813
A/Accession: T38905
A/Status: translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-474 <CON>
A/Cross-references: EMBL:Z99261; NID:g4038621; PIDN:CAB16394.1; PID:g2414656; GSPDB:GN000

A; Experimental source: strain 972h-; cosmid c56E4
C; Genetics:
A; Gene: SPDB:SPAC56E4.03
A; Map position: 1
C; Superfamily: Escherichia coli valine-pyruvate transaminase
C; Keywords: aminotransferase; phosphoprotein; pyridoxal phosphate
P: 297/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

	Query Match	4.6%; Score 95; DB 2; Length 474;	
	Best Local Similarity	22.3%; Pred. No. 1.2;	
	Matches	69; Conservative 48; Mismatches 125; Indels 68; Gaps 17;	
QY	33 SSPAGGEDRVHVLSS-----WRSGSFLGQLFSQHDPDVFYLMEPAWHVW	78	
DB	: : :: : : : :	127	
QY	79 T-TLSQGSAA TLHMAVRDL-----MRSTFL-----CDMDVFDAYMP	113	
DB	: : : : : : : : : : : : :	186	
QY	114 QSRNLSAFFNW-ATSRALCSPPACSAFPRGTISKQDVCKTCLCTROPFLAREACRSYSHV	172	
DB	: : : : : : : : : : : : :	241	
QY	173 VLKEVREFNLOV-LY---PLLSDPALNLRIVHLVRDPRAVLRSREAAAPILARDN--GI	225	
DB	: : : : : : : : : : : : :	299	
QY	226 VLGTNKGKWEADPHRLIREVCRRSHVRIAEEAATLKPPPFLRG-RYRLVRF--EDLAREPL	282	
DB	: : : : : : : : : : : : :	353	
QY	283 AETRALYAFT 292		
DB	:		
QY	354 KHIR-YSYT 361		
DB	:		

RESULT 10

S78246

endopeptidase Clp (EC 3.4.21.-) ATP-binding chain clpC [similarity] - Odont

N/Alternate names: ATP-dependent Clp proteinase regulatory chain; caseinoly

N/Contains: adenosinetriphosphatase (EC 3.6.1.3)

C/Species: chloroplast Odontella sinensis

C/Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 19-Jan-2001

C/Accession: S78246

R/Kowalik, K.V.; Stoebe, B.; Schaffran, I.; Kroth-Pancic, P.; Freier, U.

Plant Mol. Biol. Rep. 13, 336-342, 1995

A/Title: The Chloroplast Genome of a chlorophyll a+c- containing Alga, Odon

A/Reference number: S78238

A/Accession: S78246

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-885 <KOW>

A/Cross-references: EMBL:Z67753; NID:g1185127; PIDN:CAA91619.1; PID:g118513

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, Nov

C/Genetics:

A/Gene: clpC

A/Genome: chloroplast

C/Function:

A/Description: allows clpp to hydrolyze polypeptides and proteins, probably

e activity; ATP hydrolysis is required for Clp hydrolysis of proteins but n

C/Superfamily: endopeptidase Clp ATP-binding chain

C/Keywords: ATP; chloroplast; hydrolase; molecular chaperone; nucleotide bi

F/218-225/Region: nucleotide-binding motif A (P-loop)

F/285-290/Region: nucleotide-binding motif B

F/560-567/Region: nucleotide-binding motif A (P-loop)

F/628-633/Region: nucleotide-binding motif B

F/224/Binding site: ATP (Lys) #status predicted

F/566/Binding site: ATP (Lys) #status predicted

Query Match 4.6% Score 95 DB 1 Length 885

Query Match 4.68; Score 95; DB 1; Length 885;

Best Local Similarity 21.0%; Pred. No. 2.8;									
Matches 89; Conservative 59; Mismatches 161; Indels 114; Gaps 18;									
QY	5	RFSSKTVVLLAQ-----TCLLLFII-SRPGSPAGGEDRVHVLVLSWR	51						
Db	4	KFTEGAIKVIMLSQEEARRMGHNVGTEQLLLGIQRHGIGARALKQKV---TLKKAR	60						
QY	52	-----SGSSFGLQFSQHPDVVYLMPEPAHWVTTLSQGSAAFLHMAVRDLMRS--LF	101						
Db	61	REIELYIGRTGFVASEIPFTPRAKRVLEMVAVHEGKDLGQNEFVGTETHILLALISESGVA	120						
QY	102	LCDMDVFDAYMPQSRNLSAFFNWATSRALCSPACSAFPRGTSISKQDVCKTLCTRQPFSL	161						
Db	121	MRTLDKLGVNIPKLRNLILMYI-----EENQEEILRPLTQAEKFL	161						
QY	162	AREACKS-----YSHVVLKEVREFNLQVLYPLLSDPALNL-RIVHLVRDPRAVLRSR	213						
Db	162	EREKKGSTPTLDEYSENISKFAVDGKL-----DPVIGRDKEIHEV---IKVLARR	210						
QY	214	AAGPILARDNGIVLTNGKWVEADPHLRLLIREVCRSHVRIAEAAATLKPPPFELRGYRLVR	273						
Db	211	KNNPVLIGPGV-----GKTAVAEGLAQLI-----IAB-----KAPDFLDGNL----	248						
QY	274	FEDLAREPLAEIRALYAFTGLTLTPOLEAWIHNITHGSGIGKPIEAFHT-----SSRNAR	328						
Db	249	---LMALDLGSILAGTKYRG-EFEERIKRIVEEVQNSAILVIDEIHITLVGAGAECAV	304						
QY	329	NVSQAWRHALPFTKILRVQEVCAQALQLLYR-----PVYSADQQRDLTLD	374						
Db	305	DAANILKPALARGKFR-----CIGATTIDEYRKYIERDPALERRFQPVHVKEPTVGVTIE	359						
QY	375	LVL 377							
Db	360	ILL 362							
RESULT 11									
VGWJBV									
peplomer glycoprotein precursor - Berne virus (strain P138/72)									
C;Species: Berne virus									
C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jul-1999									
C;Accession: A36759									
R;Snijder, E.J.; Den Boon, J.A.; Spaan, W.J.M.; Weiss, M.; Horzinek, M.C.									
Virology 178, 355-363, 1990									
A;Title: Primary structure and post-translational processing of the Berne virus peplomer									
A;Reference number: A36759; MUID:91020973									
A;Accession: A36759									
A;Molecule type: genomic RNA									
A;Residues: 1-1581 <SNI>									
A;Cross-references: GB:X52506; NID:g62059; PIDN:CAA36748.1; PID:g62060									
C;Genetics:									
A;Gene: P									
C;Superfamily: Berne virus peplomer glycoprotein									
C;Keywords: glycoprotein; transmembrane protein									
F:1-19/Domain: signal sequence #status predicted <SIG>									
F:20-1581/Product: peplomer glycoprotein #status predicted <PGP>									
F:1547-1572/Domain: transmembrane #status predicted <TMN>									
F:25,310,384,494,574,935,969,1267,1297,1327,1385,1389,1428,1431,1438,1483,1487,1495,1515									
Query Match 4.6%; Score 94.5; DB 1; Length 1581;									
Best Local Similarity 17.6%; Pred. No. 6.7;									
Matches 81; Conservative 65; Mismatches 135; Indels 179; Gaps 19;									
QY	2	WLPRFSSKTVVLLLAQTCLLLFIISRPGPSAGGEDRVHVLVLSWRSGSFLGQLF	61						
Db	259	WVAFQNKATAVILPSELIVPVAQKVR-----RLG 289							
QY	62	SQHPDVFYLMPEPAHWVTTLSQGSAAFLHMAVRDLMRSIFLCDMDVFDAYMPQSRNLSAF	121						
Db	290	VNTPDYFWLVKQAYH-----YLSQA-NLSP- 313							
QY	122	FNWATSRALCSPACSAFPRGTSISKQDVCKTLCTRQPFSLAREACRSYSHVV-LKEVREF	180						

Db	314	-NYALFSALCN-----SLYQQSATLSLTCFGSPFFVAQEC---YNNALYLPDAVFT	360						
QY	181	NL-----QVLYPL-----LSDPALNL-----RIVHLVRDPRAVL----	209						
Db	361	TLFSTLFSWDYQINYPNLNOVLTONETFLQLPATNYQGOTLSQGRMLNLFKDAIVFLDFFD	420						
QY	210	----RSREAAGP----ILARDNGIVLTNGKWVEADPHLRLLIREVCRSHVRIAEAAATLKP	261						
Db	421	TKFYRTNDAPSSDIFVVVARQAQLIRYGNPRIEQINGYFQV---KCSSNI-----ISTLEP	473						
QY	262	PPFLRGYRLVRFEDLAREPLAEIRALYAFT-----	292						
Db	474	HP--AGVIMIAHHSMWSVAARNSTSFYCVTHSLTTFGKLDISTSWFFHILALPSGPVSQ	531						
QY	293	-----GLTLIPQLEAWIHNIT-----HGSGIGKPIEAFHTSSRNARNS	331						
Db	532	VSMPLLSTAAGVVMHPMIEHWIPLTLTAQSQYQPSFFNIGINKTI-TLTTQLQAYAOVY	590						
QY	332	QAWRHALPFTKILRVQEVCAQALQLLYRFPVYSADQQORDL	371						
Db	591	TAWFLSVIYVRLPEARRLTLG-VQLVFFIQAALLSIKQADL	629						
RESULT 12									
A70841									
probable oxalyl-CoA decarboxylase - Mycobacterium tuberculosis (strain H37RV)									
C;Species: Mycobacterium tuberculosis									
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000									
C;Accession: A70841									
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon									
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,									
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.									
Nature 393, 537-544, 1998									
A;Authors: Sqaures, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.									
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno									
A;Reference number: A70500; MUID:98295987									
A;Accession: A70841									
A;Status: preliminary; nucleic acid sequence not shown; translation not shown									
A;Molecule type: DNA									
A;Residues: 1-582 <COL>									
A;Cross-references: GB:AL021926; GB:AL123456; NID:g3261520; PIDN:CAAL7312.1; PID:g290									
A;Experimental source: strain H37Rv									
C;Genetics:									
A;Gene: oxca									
C;Superfamily: oxalyl-CoA decarboxylase; thiamin pyrophosphate-binding domain homolog									
Query Match 4.6%; Score 94; DB 2; Length 582;									
Best Local Similarity 21.4%; Pred. No. 2;									
Matches 55; Conservative 44; Mismatches 98; Indels 60; Gaps 11;									
QY	152	TLCTRQPFSLAREACRSYSHVVLKEVFFNLQVLYPLLSDP-----ALNLRIVHLV	202						
Db	2	TTRSASPCTVLIDGC-----HLVVDALKANDVDVTIYGVVGIPITDLARAAQAAGIRYIGF-	56						
QY	203	RDPRAVLRSREAAGPILARDNGIVLTNGK-WVEADPHLRLLIREVCRSHVRIAEAAATLKP	261						
Db	57	RHEASAGNAAAAGFLTARP-GVCLTTSGGFGLNLPALANATTNCFPMIQISGSSS-RP	114						
QY	262	-----PPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPQLEAW	303						
Db	115	MVDLQRGDYQDLDLQNAARPFVKAAAYRIGVQDQIGRGVARAIR-----TATSGRPGG	166						
QY	304	IHNITHGSGIGKPIEAFHTSSRNARNSQAWRHALPFTKILRVQEVCAQALQLLYRFPVY	363						
Db	167	VYLDIPGDVLGQAVEASAAS-----GAIWRPVDPAIRLLPAPEAIDRALDVL-----	213						
QY	364	SADQQRDLTLDLVLPRG	380						
Db	214	-AQQRPL---LVLSKG	226						

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.
OM protein - protein search, using sw model
Run on: January 31, 2002, 09:51:05 ; Search time 16.97 Seconds
(without alignments)
842.621 Million cell updates/sec
Title: US-09-593-828-8
Perfect score: 2051
Sequence: 1 MWLPRFSKTVTVLLAQIT.....LTLDLVLPRGPDHFSWASPD 390

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	540.5	26.4	458	1 C6ST_CHICK	Q92179 gallus gall
2	105	5.1	359	1 TPSE_CAEEL	Q20351 caenorhabdi
3	102.5	5.0	388	1 YM67_MYCTU	Q50695 mycobacteri
4	96	4.7	388	1 ODPT_HUMAN	P29803 homo sapien
5	96	4.7	707	1 PALL_IPOBA	P14166 ipomoea bat
6	95	4.6	885	1 CLPC_ODOSI	P49574 odontella s
7	94.5	4.6	1581	1 VGLP_BEV	P23052 berne virus
8	90.5	4.4	380	1 TPSE_CAEEL	O77081 caenorhabdi
9	89	4.3	371	1 CYB_LATCO	Q9mlj4 laticauda c
10	88.5	4.3	249	1 NODH_RHTR	P52994 rhizobium t
11	86.5	4.2	376	1 TPS2_MOUSE	O88856 mus musculu
12	86.5	4.2	377	1 TPS2_HUMAN	O60704 homo sapien
13	86.5	4.2	1608	1 RRPL_BDV	P52639 borna disea
14	85.5	4.2	221	1 YPRE_CORGL	P46584 corynebacte
15	85.5	4.2	389	1 ODPA_PIG	P29804 sus scrofa
16	85	4.1	1182	1 RTP2_TRYBG	P15594 trypanosoma
17	84.5	4.1	390	1 ODPA_MOUSE	P35486 mus musculu
18	84.5	4.1	390	1 ODPA_RAT	P26284 rattus norv
19	84.5	4.1	491	1 C2F1_HUMAN	P24903 homo sapien
20	84.5	4.1	1844	1 POLR_TYMV	P20128 turnip yell
21	83.5	4.1	390	1 ODPA_HUMAN	P08559 homo sapien
22	83.5	4.1	1844	1 POLR_TYMV	P10358 turnip yell
23	83.5	4.1	1844	1 POLR_TYMC	P28477 turnip yell
24	83	4.0	363	1 ODPA_SMIMA	P52900 sminthopsis
25	83	4.0	391	1 ODPT_MOUSE	P35487 mus musculu
26	82	4.0	746	1 PPK_STRCO	Q9kzv6 streptomyce
27	81.5	4.0	703	1 PPK_STRGR	Q9eus8 streptomyce
28	81.5	4.0	715	1 PERE_HUMAN	P11678 homo sapien
29	81.5	4.0	3011	1 POLG_HCV1	P26654 h genome po
30	81	3.9	882	1 Y890_MYCTU	Q10550 mycobacteri
31	81	3.9	972	1 MTRA_YEAST	Q99189 saccharomyc
32	80.5	3.9	704	1 DP3E_RHOCA	O68045 rhodobacter
33	80.5	3.9	976	1 EPA1_HUMAN	P21709 homo sapien

34	80.5	3.9	1942	1 Y054_HUMAN	P42694 homo sapien
35	80	3.9	247	1 NOH4_RHIME	P06237 rhizobium m
36	80	3.9	328	1 RPOA_NEIMA	Q9jr06 neisseria m
37	80	3.9	392	1 BML5_HUMAN	O95972 homo sapien
38	80	3.9	415	1 CXA3_RAT	P29414 rattus norv
39	80	3.9	416	1 CXA3_MOUSE	Q64448 mus musculu
40	80	3.9	933	1 ODOL_ECOLI	P07015 escherichia
41	79.5	3.9	559	1 MASY_MAIZE	P49081 zea mays (m
42	79	3.9	738	1 PLAK_XENLA	P30998 xenopus lae
43	78	3.8	462	1 TRPE_THETH	P05378 thermus aqu
44	77.5	3.8	370	1 TPST_HUMAN	O60507 homo sapien
45	77.5	3.8	370	1 TPST_MOUSE	O70281 mus musculu

ALIGNMENTS

RESULT 1
C6ST_CHICK
ID C6ST_CHICK STANDARD; PRT; 458 AA.
AC Q92179;
DI 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CHONDROITIN 6-SULFOTRANSFERASE (EC 2.8.2.17) (C6ST).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-44; 46-52 AND 186-192.
RC STRAIN=WHITE LEGHORN; TISSUE=Embryonic chondrocytes;
RX MEDLINE=95355490; PubMed=7629189;
RA Fukuta M., Uchimura K., Nakashima K., Kato M., Kimata K.,
RA Shinomura T., Habuchi O.;
RT "Molecular cloning and expression of chick chondrocyte chondroitin 6-
sulfotransferase.";
RL J. Biol. Chem. 270:18575-18580(1995).
CC -!- FUNCTION: CATALYZES THE SULFATION OF CHONDROITIN AND KERATAN
SULFATE.
CC -!- CATALYTIC ACTIVITY: 3'-PHOSPHOADENYLYLSULFATE + CHONDROITIN =
ADENOSINE 3',5'-BISPHOSPHATE + CHONDROITIN 6'-SULFATE.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI MEMBRANE
(BY SIMILARITY).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; D49915; BAA08655.1; --
Transferrase; Transmembrane; Signal-anchor; Glycoprotein; Golgi stack.
DOMAIN 1 23 CYTOPLASMIC (POTENTIAL).
TRANSMEM 24 37 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
DOMAIN 38 458 LUMENAL, CATALYTIC (POTENTIAL).
CARBOHYD 62 73 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 73 73 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 236 236 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 443 443 N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQUENCE 458 AA; 52253 MW; C9A3B7D0A5086F0C CRC64;

Query Match 26.4%; Score 540.5; DB 1; Length 458;
Best Local Similarity 34.3%; Pred. No. 3e-41;
Matches 124; Conservative 66; Mismatches 141; Indels 31; Gaps 8;
QY 36 AGGEDRVHVLVLSWRSGLQFLSQHPDVFYLMPEAHV--WTTLSQGSAAFLHMAV 93

Db 107 AAPEPRRHVLLMATTTRGSGSVGEFFNQGNIFLYFEPLWHIERTVTTFEPGGANAVGSAL 166
QY 94 --RDLMSIFLCMDVFDAYM---PQSRNLSAFFNWTASRALCSPACSAFPRGTISKQD 148
Db 167 VYRDVLQQLLLCDLYIIESFTSPAPEEHLTAALFRRGSSHSCLCEEPVCTPSLKKVFEKYH 226
QY 149 VCKTLCTRQPFSLAREACRSYSHVVLKEVRFENLVLYPLSDPALNLRIVHLVRDPRV 208
Db 227 CKNRRCGPNITLAAEACRRKQHMAIKTVRIQLEFLQPLAEDPRDLRIIQLVRDPRV 286
QY 209 LRSREAAGPILARDNGIVLGTNGKWVE-----ADPHRLRREVCRSHVRIAEAAATL 259
Db 287 LVSRMVA-----FSGKYESWKKWAAGEAPLQDEEVQRLGNCE-IRLSAELGL 335
QY 260 KPPPELRCRYELVRFFEDLAREPLAEIRALYAFTGLTLTPQLEAWIHNTGSGIGKPIEA 319
Db 336 RQPRWLGRYMLVRYEDVARAPLRKALEMYRFAGIHPTPQVEEWIRANTQAP---QDSNG 392
QY 320 PHTSSRNARNVQAWRHALPFTKILRVQEVACAGALQILGYRPVYSADQQRDLTLDLVLPR 379
Db 393 IYSTQKNSSEQFEKWRFSIPFKLAQVVQDACEPAMRLFGYKLASSAQELNRSLSL-LEE 451
QY 380 GP 381
Db 452 GP 453

RESULT 2
TPSB_CAEEL STANDARD; PRT; 359 AA.
AC Q20351;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PUTATIVE PROTEIN-TYROSINE SULFOTRANSFERASE (EC 2.8.2.20)
DE (TYROSYLPROTEIN SULFOTRANSFERASE) (TPST).
GN F42G9.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Taich A.;
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CATALYZES THE O-SULFATION OF TYROSINE RESIDUES WITHIN
CC ACIDIC MOTIFS OF POLYPEPTIDES (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 3'-PHOSPHOADENYLYLSULFATE + PROTEIN TYROSINE =
CC ADENOSINE 3',5'-BISPHOSPHATE + PROTEIN TYROSINE-O-SULFATE.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI MEMBRANE (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE PROTEIN SULFOTRANSFERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U00051; AAA91354.1; -.
DR WormPep; F42G9.8; CE07235.
DR Hypothetical protein; Transferase; Transmembrane; Glycoprotein;
KW Signal-anchor. 1 27 CYTOPLASMIC (POTENTIAL).
KW DOMAIN 28 48 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT TRANSMEM (POTENTIAL).
FT DOMAIN 49 359 LUMENAL, CATALYTIC (POTENTIAL).
FT SEQUENCE 359 AA; 41469 MW; 9FC2F44539B8CC53 CRC64;
SQ

Query Match 5.1%; Score 105; DB 1; Length 359;
Best Local Similarity 19.8%; Pred. No. 0.049;
Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;
QY 11 VTVLLLAQTCLLLFI-----ISRPGSPSPAGGEDRVH-----VLV 46
Db 31 IYIFIFCFTCLLIFSSIKCKKLEQKLEQLSKESLIFNEQDARHSRLLSNLEQLIFV 90
QY 47 LSSWRSGSSELQLESQHPDV-----FYLMEPAW-HVWTTLSQGSAAATLHMAVR 94
Db 91 GGVPRTGFTLRAILDADHPDVRCGGETMLDPSLFTWQAGWRNDWV---NNSGIT----- 141
QY 95 DLMRSIFLCMDVFDAYMPQSRNLSAFFNWTASRALCSPACSAFPRGTISK-QDVCKTL 153
Db 142 -----QEVED-----DAVSAFITEIVAKHSELAPRL 167
QY 154 CTROPFSLAREACRSYSHVVLKEVRFENLVLYPLSDPALNLRIVHLVRDPRVLSR- 212
Db 168 CNKDP-----YTALWLPTIR-----RLYP-----NAKFILMIRDARAVVHSMI 205
QY 213 EAAGPILARDNGIVLGTNGKWVEADPHRLRREVCRSHVRIAEAAATLKPPPELGRYRLV 272
Db 206 ERKVPVAGYNTSDEISMVQW---NOELRKMTFQCNN---APQCICK-----V 247
QY 273 RFEDLAREPLAEIRALYAFTGLTLTPQLEAWIHNTGSGIGKPI---EAFHTSS-RNA 327
Db 248 YERLIQKPAEEILRITNFDLPFSQOM-----LRHQDLIGDEVLDNQDEFSASQVKNS 301
QY 328 RNVS---QAWRHALPFTKILRVQEVACAGALQILGY-----RPVYS 364
Db 302 INTKALTSWEDCFSEETLRKLDLV-APFLGILGYDTSISKPDYS 344
RESULT 3
YM67_MYCTU STANDARD; PRT; 388 AA.
AC Q50695;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 46.1 KDA PROTEIN RV2267C.
GN RV2267C OR MT2329 OR MTCY339.43.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Mycobacterium.
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RL


```

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC DR EMBL; 277163; CAB00968.1; -.
CC DR EMBL; AE007076; AAK46611.1; -.
CC DR TIGR; MT2329; -.
CC DR TubercuList; RV2267c; -.
CC KW Hypothetical protein; Complete proteome.
CC SQ SEQUENCE 388 AA; 46062 MW; 5DED0263275A9B24 CRC64;

```

```

Query Match 5.0%; Score 102.5; DB 1; Length 388;
Best Local Similarity 22.7%; Pred. No. 0.091;
Matches 66; Conservative 37; Mismatches 99; Indels 89; Gaps 15;

```

```

QY 44 VLVSSWRSGSFLGQLF---SQH-----PDVFLMEPAWHVVTLSQGSAAATLH 90
Db 83 IFIVGHWRGTGTLHELLVVDRTGTGTGCECLAPHHFLTE----WFAPYVEFLVSKH 137
QY 91 MAVRDLMRSIFLCMDVFDAYMPQSRNLSAFFNWATSRALCSPACSAFPRTISKQDVC 150
Db 138 RAMDNMDLSL-----HHPQEDE----FVWC-MQGLPSPILTIAFP----- 172
QY 151 KTLCTROP-----FSLAREACRSYSHVVLKEVFFNQLVLY-----PILLSDPALNLR 198
Db 173 ----NRPPQVEYLDLEQVAPRELE--TWKRTLFRFVQVYFRRRTKVTILKNPHTSFRIK 226
QY 199 -----VHLVRDPRAVLRREAAGPILARDNGIVL----GTNCKWVEADPHLR 243
Db 227 VLLEVPQAKFIHIVRDPYVVPSTIHLKALYRIHGLQOPTFDGLDDKVYST--YVDLY 284
QY 244 REVCRSHVRIAEATLKPPPFELRGYRLVRFEDLAREPLAEIRALYAFGL 294
Db 285 R-----KLDEGRELVDP-----RFYELRYEDLIGDPEGQLRLYQHLGL 324

```

```

RESULT 4
ODPT_HUMAN
ID ODPT_HUMAN STANDARD; PRT; 388 AA.
AC P29803;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PYRUVATE DEHYDROGENASE E1 COMPONENT ALPHA SUBUNIT, TESTIS-SPECIFIC
DE FORM, MITOCHONDRIAL PRECURSOR (EC 1.2.4.1) (PDHE1-A TYPE II).
GN PDHA2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91065637; PubMed=2249846;
RA Dahl H.H.M., Brown R.M., Hutchison W.M., Maragos C., Brown G.K.;
RT "A testis-specific form of the human pyruvate dehydrogenase E1 alpha
RT subunit is coded for by an intronless gene on chromosome 4.";
RL Genomics 8:225-232(1990).
CC -!- FUNCTION: THE PYRUVATE DEHYDROGENASE COMPLEX CATALYZES THE OVERALL
CC CONVERSION OF PYRUVATE TO ACETYL-COA & CO(2). IT CONTAINS MULTIPLE
CC COPIES OF THREE ENZYMATIC COMPONENTS: PYRUVATE DEHYDROGENASE (E1),
CC DIHYDROLIPOAMIDE ACETYLTRANSFERASE (E2) & LIPOAMIDE DEHYDROGENASE
CC (E3).
CC -!- CATALYTIC ACTIVITY: PYRUVATE + LIPOAMIDE = S-ACETYL-DIHYDRO-
CC LIPOAMIDE + CO(2).
CC -!- COFACTOR: THIAMINE PYROPHOSPHATE.
CC -!- ENZYME REGULATION: E1 ACTIVITY IS REGULATED BY PHOSPHORYLATION
CC (INACTIVATION) AND DEPHOSPHORYLATION (ACTIVATION) OF THE ALPHA

```

```

CC SUBUNIT.
CC -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA SUBUNITS.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -!- TISSUE SPECIFICITY: TESTIS. EXPRESSED IN POSTMEIOTIC SPERMATOGENIC
CC CELLS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC -----
CC DR EMBL; M86808; AAA60232.1; -.
CC DR PIR; A37104; DEHUP1.
CC DR MIM; I79061; -.
CC DR InterPro; IPR001017; E1_dh.
CC DR Pfam; PF00676; E1_dehydrog; 1.
CC KW Glycolysis; Oxidoreductase; Flavoprotein; Thiamine pyrophosphate;
CC KW Phosphorylation; Mitochondrion; Transit peptide; Multigene family.
CC FT TRANSIT 1 27 MITOCHONDRION (BY SIMILARITY).
CC FT CHAIN 28 388 PYRUVATE DEHYDROGENASE E1 COMPONENT
CC FT MOD_RES 230 230 ALPHA SUBUNIT, TESTIS-SPECIFIC FORM.
CC FT MOD_RES 291 291 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 298 298 PHOSPHORYLATION (BY SIMILARITY).
CC SQ SEQUENCE 388 AA; 42933 MW; 075B6CFF6DC73CC5 CRC64;

```

```

Query Match 4.7%; Score 96; DB 1; Length 388;
Best Local Similarity 23.2%; Pred. No. 0.35;
Matches 45; Conservative 25; Mismatches 82; Indels 42; Gaps 7;

```

```

QY 82 SQGS-AATLHMAVRDLMRSIFLCMDVFDAYMPQSRNLSAFFNWATSRALCSPACSAFP 140
Db 198 NQOQIAEAFNMAALWKLPCVFICENNLYG-----MGSTERAASP---DYK 242
QY 141 RGTI---SKQDVCKTLCTROPFSLAREACRSYSHVVLKEVFFNQLVLYPLSDPALNLR 197
Db 243 RGNFIPGLKVDGMDVLCVREATKFAANYCRSGKGPILMELQTYRYH--GHSMSDPGVSYR 300
QY 198 I-----VHLVRDPRAVLRREAAGPILARDNGIVLGTNGKWEADPHLRILIREVCRSHV 251
Db 301 TREIQEVRSKRDPPIILQDRM-----VNSKLATVEELKEIGAEVRKEID 345
QY 252 RIAERAATLKPPPF 265
Db 346 DAAQFATTDPEPHL 359

```

```

RESULT 5
PALL_IPOBA
ID PALL_IPOBA STANDARD; PRT; 707 AA.
AC P14156;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PHENYLALANINE AMMONIA-LYASE (EC 4.3.1.5).
GN PAL.
OS Ipomoea batatas (Sweet potato) (Batate).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Convolvulaceae; Ipomoea.
OX NCBI_TaxID=4120;
RN [1]
RP SEQUENCE FROM N.A.
RA Tanaka Y., Matsuoka M., Yamanoto N., Ohashi Y., Kano-Murakami Y.,
RA Ozeki Y.;
RT "Structure and characterization of a cDNA clone for phenylalanine
RT ammonia-lyase from cut-injured roots of sweet potato.";
RL Plant Physiol. 90:1403-1407(1989).
CC -!- FUNCTION: THIS IS A KEY ENZYME OF PLANT METABOLISM CATALYZING THE

```



```

OC Coronaviridae; Torovirus.
OX NCBI_TaxID=111156;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISOLATE P138/72;
RX MEDLINE=91020973; PubMed=2219698;
RA Snijder E.J., den Boon J.A., Spaan W.J.M., Weiss M., Horzinek M.C.;
RT "Primary structure and post-translational processing of the Berne
RL virus peplomer protein.";
CC Virology 178:355-363(1990).
CC -!- FUNCTION: THE PEPLIMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52506; CAA36748.1; -.
DR PIR; A36759; VGWJBV.
KW Glycoprotein; Envelope protein; Transmembrane; Signal.
FT SIGNAL 1 19
FT CHAIN 20 1581 PEPLIMER GLYCOPROTEIN.
FT TRANSMEM 1547 1572 POTENTIAL.
FT CARBOHYD 25 25 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 384 384 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 574 574 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 935 935 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 969 969 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1267 1267 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1297 1297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1385 1385 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1389 1389 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1428 1428 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1431 1431 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1438 1438 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1483 1483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1487 1487 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1495 1495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1515 1515 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1581 AA; 178332 MW; 00D91B41837AC769 CRC64;

Query Match 4.6%; Score 94.5; DB 1; Length 1581;
Best Local Similarity 17.6%; Pred. No. 2.9;
Matches 81; Conservative 65; Mismatches 135; Indels 179; Gaps 19;

QY 2 WLPFRSSKTVTVLLAQITCLLLFIISRPSPAGGEDRVHVLVLSWRSFGSLGQLF 61
Db 259 WVVAFQNKATAVILPSELIVPVAQKVR-----RLG 289
QY 62 SQHPDVFFYLMPEAWHVWTTLSQGSAAIHLMAVRDLMRSLFLCDMDVFDAYMPQSRNLSAF 121
Db 290 VNTPDYFWLVKQAYH-----YLSQA-NLSP- 313
QY 122 FNWATSRALCPSPACSAFPRGTISKQDVCKTLCRTPFSLAREACRSYSHV-LKEVRFF 180
Db 314 -NYALFSALCN-----SLYQQSATLSTLCFSPFFVAQEC---YNNALVLPDAVFT 360
QY 181 NL-----QVLYPL-----LSDPALNL-----RIVHLVRDPAVL----- 209
Db 361 TLFSTLFSWDYQINYPNLNQVLTONETFLQLPATNYQGQTLTSGQRMNLNFKDAIVFLDFFD 420
QY 210 ----RSREAAGP-----ILARDNGIVLGTNGKVVVEADPHRLRIREVCRSHVRIAEAAATLKP 261
Db 421 TKFYRTNDAPSSDTFVVVARQAQLIRYGNFRIEQINGYFQV---KCSSNI-----ISTLEP 473
QY 262 PFFLRGYRLVRFEDLAREPLAETRALYAPT----- 292
```

```

Db 474 HP--AGVIMIAHHSMWSVAARNSTSFYCVTHSLTTEGKLDISISWFFHTLALPSPVVSQ 531
QY 293 -----GLTLTPQLEAWIHNI-----HGSGIGKPIEAFHTSSRNARNVS 331
Db 532 VSMPLLLSTAAGVYMHPIEHWIPLLTLTAOSQYQPSFFNIGINKTI-TLTTQLQAYAQVY 590
QY 332 QAWRHALPFTKILRVQEVCAQALQLLGYRVPVYSADQQQDL 371
Db 591 TAWFLSVIYVRLPEARRLTIG-VQLVPFIQALLSIKQADL 629

RESULT 8
TPSA_CAEEL
ID TPSA_CAEEL STANDARD; PRT; 380 AA.
AC G77081; Q9NEW9;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN-TYROSINE SULFOTRANSFERASE A (EC 2.8.2.20) (TYROSYLPROTEIN
DE SULFOTRANSFERASE-A) (TPST-A).
GN Y111B2A.15.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=98406128; PubMed=9733778;
RA Ouyang Y.-B., Moore K.L.;
RT "Molecular cloning and expression of human and mouse tyrosylprotein
RT sulfotransferase-2 and a tyrosylprotein sulfotransferase homologue in
RT Caenorhabditis elegans.";
RL J. Biol. Chem. 273:24770-24774(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Sulston J.E.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Durbin R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CATALYZES THE O-SULFATION OF TYROSINE RESIDUES WITHIN
CC ACIDIC MOTIFS OF POLYPEPTIDES.
CC -!- CATALYTIC ACTIVITY: 3'-PHOSPHOADENYLYLSULFATE + PROTEIN TYROSINE =
CC ADENOSINE 3',5'-BISPHOSPHATE + PROTEIN TYROSINE-O-SULFATE.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI MEMBRANE (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE PROTEIN SULFOTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF049709; AAC36062.1; -.
DR EMBL; AL132904; CAC35844.1; -.
DR WormPep; Y111B2A.15; CE26632.
KW Transferase; Transmembrane; Glycoprotein; Signal-anchor.
FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 7 27 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT DOMAIN 28 380 (POTENTIAL).
FT CARBOHYD 66 66 LUMENAL, CATALYTIC (POTENTIAL).
SQ SEQUENCE 380 AA; 43313 MW; FF709BF00FEDC95 CRC64;

Query Match 4.4%; Score 90.5; DB 1; Length 380;
Best Local Similarity 19.3%; Pred. No. 1.1;
```


us-09-593-828-8.rsp

Fri Feb 1 20:21:12 2002

Search completed: January 31, 2002, 09:51:07
Job time: 325 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:45:09 ; Search time 42.22 Seconds
(without alignments)
684.239 Million cell updates/sec

Title: US-09-593-828-8
Perfect score: 2051
Sequence: 1 MWLPRFSSKTVTVLLAQTT.....LTLDLVLPGRPDHFSWASPD 390

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result NO.	Score	Query Match	Length	ID	Description
1	2051	100.0	390	22 AAY72639	Human glycosyl sul
2	1865.5	91.0	418	21 AAB41947	Human ORFX ORF1711
3	1729.5	84.3	395	22 AAY72640	Human glycosyl sul
4	1536.5	74.9	395	22 AAY72638	Mouse glycosyl sul
5	1008	49.1	386	20 AAY39918	Human glycosyl sul
6	970	47.3	386	21 AAY79219	Human transferase
7	958.5	46.7	388	20 AAY39919	Mouse glycosyl sul
8	665	32.4	483	20 AAY31656	Mouse N-acetylgluc
9	659.5	32.2	530	22 AAB95367	Human protein sequ
10	656.5	32.0	484	20 AAY31657	Human N-acetylgluc
11	577.5	28.2	479	19 AAW52863	Glycosaminoglycan

12	540.5	26.4	458	18 AAW06480	Chick chondroitin
13	513.5	25.0	411	19 AAW61100	Keratan sulphate 6
14	119.5	5.8	596	22 AAY72641	Human glycosyl sul
15	119.5	5.8	1222	22 AAY72642	Human glycosyl sul
16	112.5	5.5	1207	22 AAY72643	Mouse glycosyl sul
17	105	5.1	359	20 AAY06628	C. elegans tyrosyl
18	105	5.1	359	21 AAY84309	A tyrosylprotein s
19	99.5	4.9	928	22 AAE05075	Drosophila melanog
20	90.5	4.4	380	20 AAY06627	C. elegans tyrosyl
21	90.5	4.4	380	21 AAY84308	A tyrosylprotein s
22	86.5	4.2	330	21 AAB34913	Gene 7 human secre
23	86.5	4.2	330	21 AAB34914	Human secreted pro
24	86.5	4.2	376	20 AAY06626	Mouse tyrosylprote
25	86.5	4.2	376	21 AAY84307	A murine tyrosylpr
26	86.5	4.2	377	20 AAY06625	Human tyrosylprote
27	86.5	4.2	377	21 AAY84306	A human tyrosylpro
28	86.5	4.2	377	21 AAY69421	Amino acid sequenc
29	86.5	4.2	566	15 AAR55207	Human zona pelluci
30	86.5	4.2	566	20 AAY42479	Human zona pelluci
31	86.5	4.2	566	20 AAW81816	Human ZPB protein.
32	86.5	4.2	566	21 AAY82214	Human zona pelluci
33	86.5	4.2	566	21 AAY52688	Human oocyte zona
34	86.5	4.2	566	21 AAY52179	Human zona pelluci
35	86.5	4.2	566	21 AAY52983	Human zona pelluci
36	86.5	4.2	1608	17 AAR98619	Human zona pelluci
37	86.5	4.2	1711	17 AAR98605	Borna disease viru
38	86	4.2	1055	22 AAM39198	Borna disease viru
39	85.5	4.2	366	12 AAR15475	Human polypeptide
40	85.5	4.2	467	22 AAR88483	CHDH. Nocardia sp
41	84.5	4.1	484	22 AAB76522	Human membrane or
42	84.5	4.1	491	22 AAG90675	Corynebacterium gl
43	84.5	4.1	491	22 AAB45789	C glutamicum prote
44	84	4.1	568	21 AAB32533	C. glutamicum accD
45	83.5	4.1	977	18 AAW19258	S. lavendulae Mmct
					Embryonic stem cel

ALIGNMENTS

RESULT 1	
AAY72639	
ID	AAY72639 standard; Protein; 390 AA.
XX	
AC	AAY72639;
XX	
DT	02-MAY-2001 (first entry)
XX	
DE	Human glycosyl sulfotransferase-4alpha (GST-4alpha).
XX	
KW	Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation;
KW	systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW	polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW	glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis;
KW	Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW	demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW	myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW	asthma; hypersensitivity; rheumatic fever; tissue rejection;
XX	chromosome 16q23.1.
OS	Homo sapiens.
XX	
PN	WO200106015-A1.
XX	
PD	25-JAN-2001.
XX	
PF	19-JUL-2000; 2000WO-US19741.
XX	
PR	20-JUL-1999; 99US-0144694.
PR	13-JUL-2000; 2000US-0593828.
XX	
PA	(REGC) UNIV CALIFORNIA.
XX	

PI Rosen SD, Lee JK, Hemmerich S;
XX WPI; 2001-138471/14.
DR N-PSDB; AAD02697, AAD02698, AAD02699.
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT diagnostic and therapeutic agent screening applications -
XX Claim 3; Fig 1; 128pp; English.
PS
XX The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
CC alpha). GST-4 gene is found on chromosome 16q23.1.
CC GST is a type 2 membrane protein useful for inhibiting a binding event
CC between a selectin and a selectin ligand, which comprises contacting the
CC selectin with a non-sulphated selectin ligand, GST and a small molecular
CC agent that inhibits the sulphation activity of GST. GST is also useful
CC in inhibiting a selectin mediated binding event. GST is useful in gene
CC therapy to treat disorders such as acute or chronic inflammation,
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC disease, Grave's disease, adenitis, hypoparathyroidism, pernicious
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation.
CC Note: The present sequence is also shown in sequence listing (page
CC no: 56) but lacks four nucleotides at its 3' end.
XX
SQ Sequence 390 AA;

Query Match 100.0%; Score 2051; DB 22; Length 390;
Best Local Similarity 100.0%; Pred. No. 1.9e-216;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MWLPRFSSKTVTVLLLAQTCLLLFIISRPSPSPAGGEDRVHVLVLSWRSGSFLGQL 60
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1 mwlpfsskttvtvlllaqttclllfiisrpgpspaggedrvhvlvsswrgsflgql 60
QY 61 FSQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLMRSIFCLDMDFDAYMPQSRNLSA 120
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
61 fsqhpdvfylnepawhvwttlsqgsaatlhnavrdlmrsifclmdvfdaympqsrnlsa 120
QY 121 FFWATSRALCSPACAPPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRFF 180
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
121 ffwatrsalcspacsafrgtiskdqvcktlctrqpfslareacrsyshvvlkevrf 180
QY 181 NLQVLYPLLSDPALNLRIVHLVRDPRAVLRSREAGPILARDNGIVLGTNGKWWVADPHL 240
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
181 nlqvlypllsdpalnrlrivhlvrdrpravlrsreagpillardngivlgtngkwvreadphl 240
QY 241 RLIREVCRSHVRIAEAAATLKPPPFRLGRYRLVRFEDLAREPLAEIRALYAFTGLTLPQL 300
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
241 rlirevcrshvriaeaatlkpfpflrgyrlvrfedlareplaeiralyaftgltlpql 300
QY 301 EAWIHNITHGSGIGKPIEAFTSSRNARNVSWAQRHALPFTKILRVOEVCAGALQLLGYR 360
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
301 eawihnithgsgigkpieaftssrnarnvswaqrhalpftkrlrvqevcagalqllyr 360
QY 361 PVYSADQQORDLTLDLVLPRGPDHFSWASPD 390
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
361 pvsadqqrldtldlvlprgpdhfswaspd 390

RESULT 2
AAB41947
ID AAB41947 standard; Protein; 418 AA.
XX
AC AAB41947;
XX
DT 08-FEB-2001 (first entry)

XX DE
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX Homo sapiens.
OS
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach M;
PI
XX WPI; 2000-602362/57.
DR N-PSDB; AAC76156.
DR
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 2599-2600; 5507pp; English.
PS
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 418 AA;

Query Match 91.0%; Score 1865.5; DB 21; Length 418;
Best Local Similarity 91.0%; Pred. No. 5e-196;
Matches 356; Conservative 9; Mismatches 25; Indels 1; Gaps 1;

QY 1 MWLPRFSSKTVTVLLLAQTTCLLLFIISRPSPAGGEDRVHVLVLSWSRSGSFLGQL 60
Db 28 mwlpfsskttvlllaqtclllfiisrpgppspaggedrvhvlvlsrsgssflgql 87
QY 61 FSQHPDVFYLMPEAWHVWTTLSQGSAAATLHMAVRDLMRSLFCLDMDFDAYM-PQSRNLS 119
Db 88 fsqhpdvfylnepawhvwttlsqgsaatlhmvdrldmrsifclmdvdaymepgrqs 147
QY 120 AFFNWATSRALCSPACSAFPRGTISKQDVCKTLCTRPQPSFLAREACRSYSHVVLKEVRF 179
Db 148 slfqwensralcsapacdliipqdeiiprahcrllcsqgpfevvekacrsyshvvlkevrf 207
QY 180 FNQLVLYPLSDPALNLRIVHLVRDPRAVLRSRREAAGPILARDNGIVLGTNGKWEADPH 239
Db 208 fnlqslppllkpslnlhivhlvrdrvavlrreagpilarndngivlgtngkweadph 267
QY 240 LRLIREVCRSHVRIAEAAATLKPPPFRLGRYRLVRFEDLAREPLAEIRALYAFTGLTLTPQ 299
Db 268 lrlirevcrshvriaeaatlkpfpflrgryrlvrfedlareplaeiralyaftgltltpq 327
QY 300 LEAWIHNTHSGGIGKPIEAFHTSSRNARNVSQAWRHALPFTKILRVQEVCGALQLLGY 359
Db 328 leawihnthsggigkpieafhtssrnarnvsqawrhalfpftkilorvgevcagalqlly 387
QY 360 RPVYSADQQRDLTLDLVLPRGPDHFSWASPD 390
Db 388 rpvysadqqrldtldlvprrgpdhfswaspd 418

RESULT 3

AAAY72640
ID AAY72640 standard; Protein; 395 AA.
XX
AC AAY72640;

02-MAY-2001 (first entry)

Human glycosyl sulfotransferase-4beta (GST-4beta).

Human; glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation; systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes; polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis; glomerulonephritis; myasthenia gravis; Sjogren's syndrome; Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia; demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis; myocarditis; adult respiratory distress syndrome; eczema; psoriasis; asthma; hypersensitivity; rheumatic fever; tissue rejection; chromosome 16q23.1.

Homo sapiens.

WO200106015-A1.

25-JAN-2001.

19-JUL-2000; 2000WO-US19741.

20-JUL-1999; 99US-0144694.

13-JUL-2000; 2000US-0593828.

(REGC) UNIV CALIFORNIA.

Rosen SD, Lee JK, Hemmerich S;

WPI; 2001-138471/14.

N-PSDB; AAD02697, AAD02700.

New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for diagnostic and therapeutic agent screening applications -

Claim 3; Fig 4B; 128pp; English.

CC The present sequence is human glycosyl sulfotransferase-4beta (GST-4
CC beta). GST-4 gene is found on chromosome 16q23.1.
CC GST is a type 2 membrane protein useful for inhibiting a binding event
CC between a selectin and a selectin ligand, which comprises contacting the
CC selectin with a non-sulphated selectin ligand, GST and a small molecular
CC agent that inhibits the sulphation activity of GST. GST is also useful
CC in inhibiting a selectin mediated binding event. GST is useful in gene
CC therapy to treat disorders such as acute or chronic inflammation,
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation.

XX Sequence 395 AA;

Query Match 84.3%; Score 1729.5; DB 22; Length 395;
Best Local Similarity 85.8%; Pred. No. 4e-181;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

QY 1 MWLPRFSSKTVTVLLLAQTTCLLLFIISRPSPAGGEDRVHVLVLSWSRSGSFLGQL 60
Db 1 mwlpfsskttvlllaqtclllfiisrpgppspaggedrvhvlvlsrsgssflgql 59

QY 61 FSQHPDVFYLMPEAWHVWTTLSQGSAAATLHMAVRDLMRSLFCLDMDFDAYM-PQSRNLSA 120
Db 60 fsqhpdvfylnepawhvwttlsqgsaatlhmvdrldmrsifclmdvdaylpwrrnlsd 119

QY 121 FFWATSRALCSPACSAFPRGTISKQDVCKTLCTRPQPSFLAREACRSYSHVVLKEVRF 180
Db 120 lfwavsralscspacsaafprgaisseavckplcarqstlareacrsyshvvlkevrf 179

QY 181 NLOVLYPLSDPALNLRIVHLVRDPRAVLRSRREAAGPILARDNGIVLGTNGKWEADPHL 240
Db 180 nlqvlyplsdpalnlrivilvrdrvavlrreagtakalarndngivlgtngtweadpql 239

QY 241 RLIREVCRSHVRIAEAAATLKPPPFRLGRYRLVRFEDLAREPLAEIRALYAFTGLTLTPQL 300
Db 240 rvvrevcrshvriaeaatlkpfpflrgryrlvrfedlareplaeiralyaftglstltpql 299

QY 301 EAWIHNTHSGGIGKPIEAFHTSSRNARNVSQAWRHALPFTKILRVQEVCGALQLLGYR 360
Db 300 eawihnthsggpggarreakftssrnalnvsgawrhalfakirrvqelcagalqllyr 359

QY 361 PVSADQQRDLTLDLVLPRGPDHFSWAS 388

Db 360 pvsadeqrnlaldlvprrgngftwas 387

RESULT 4

AAAY72638

ID AAY72638 standard; Protein; 395 AA.

XX

AC AAY72638;

02-MAY-2001 (first entry)

Mouse glycosyl sulfotransferase-4 (GST-4).

Mouse; glycosyl sulfotransferase-4; GST-4; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation; systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes; polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis; glomerulonephritis; myasthenia gravis; Sjogren's syndrome; Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia; demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis; myocarditis; adult respiratory distress syndrome; eczema; psoriasis; asthma; hypersensitivity; rheumatic fever; tissue rejection;

KW chromosome 8E1.
XX Mus musculus.
OS
XX WO200106015-A1.
PN
XX 25-JAN-2001.
PD
XX 19-JUL-2000; 2000WO-US19741.
XX
PF 20-JUL-1999; 99US-0144694.
PR 13-JUL-2000; 2000US-0593828.
XX
XX (REGC) UNIV CALIFORNIA.
PA
XX Rosen SD, Lee JK, Hemmerich S;
PI WPI; 2001-138471/14.
XX N-PSDB; AAD02696.
DR
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT diagnostic and therapeutic agent screening applications
PT
XX
XX Claim 3; Fig 2; 128pp; English.
XX
XX The present sequence is mouse glycosyl sulfotransferase-4 (GST-4). GST-4
CC gene is found on chromosome 8E1.
CC GST is a type 2 membrane protein useful for inhibiting a binding event
CC between a selectin and a selectin ligand, which comprises contacting the
CC selectin with a non-sulphated selectin ligand, GST and a small molecular
CC agent that inhibits the sulphation activity of GST. GST is also useful
CC in inhibiting a selectin mediated binding event. GST is useful in gene
CC therapy to treat disorders such as acute or chronic inflammation,
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation.
XX
XX Sequence 395 AA;
SQ
Query Match 74.9%; Score 1536.5; DB 22; Length 395;
Best Local Similarity 76.0%; Pred. No. 6.3e-160;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;
QY 1 MWLPRFSSKTVTVLLLAQTTCLLLFIIISRPSPAGGEDRVHVLVSSWRSGSFLGQL 60
Db 1 mrlprfsstvmllmvqtg-ilvflvsrgvpspaglgerhvhvlvsswrsgsfvgql 59
QY 61 FSQHDPDVFYLMPPAWHVVTTLSQGSAAATLHMAVRDLMRSLFCLCDMDVFDAYMPQSRNLSA 120
Db 60 fsqhdpdvfylmepawhvwtdtllsqgsapallhmvdrdlrsvflcdmdvfdaylpwrnisd 119
QY 121 FFWWATSRALCSPACSAFPRTGTSKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRFF 180
Db 120 lfgwavsralcspvceafagnisseevckpicatrfpfglaeacssyshvvlkevrf 179
QY 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSRERAGPILARDNGIVLGTNGKWEADPHL 240
Db 180 nlqvlyplsdpalnhrivhlvrdrpravlrsreqtakalardngivlgtngtwveadprl 239
QY 241 RLIREVCRSHVRIAEATLKPPPLRGYRLVRFEEDLAREPLAEIRALYAFTGLTLPQL 300
Db 240 rvvnevcrshvriaealhkppplfldryrlvryediardpltvirelyaftgltlpql 299
QY 301 EAWIHNITHSGGIGKPIEAFHTSSRNARNVQAWRHALPFTKILRVQEVCAQALQLLYR 360
Db 300 qtwihnthsgpggarreaftktsrdalsvsawrhtlpfakirrvqelcggalqllyr 359

QY 361 PVYSADQQRDLTLDLVLPRGPDHFSWAS 388
Db 360 svhselegrdlslldllprgmdsfkwas 387

RESULT 5
AAAY39918
ID AAY39918 standard; Protein; 386 AA.
XX
AC AAY39918;
XX
DT 08-DEC-1999 (first entry)
XX
DE Human glycosyl sulfotransferase-3 protein sequence.
XX
KW Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
KW selectin binding interaction; inflammation; lymphocyte homing; human;
KW secondary lymph organ.
XX
OS Homo sapiens.
XX
PN WO9949018-A1.
XX
PD 30-SEP-1999.
XX
PF 26-FEB-1999; 99WO-US04316.
XX
PR 20-MAR-1998; 98US-0045284.
PR 12-NOV-1998; 98US-0190911.
XX
XX (REGC) UNIV CALIFORNIA.
PA (SYNT) SYNTAX USA INC.
XX
XX Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
PI WPI; 1999-580442/49.
XX N-PSDB; AAZ20792.
XX
XX Human and murine glycosyl sulfotransferase 3 and related
PT polynucleotides
XX
PS Claim 2; Fig 1; 59pp; English.
XX
CC This sequence is the human glycosyl sulfotransferase-3 (GST-3) of
CC the invention. The nucleic acid sequences, probes and primers derived
CC from these, proteins and antibodies are useful in detecting homologues.
CC The sequences, antibodies and methods are useful in the diagnosis and
CC treatment of diseases associated with selectin binding interactions,
CC including conditions associated with or resulting from the homing of
CC leukocytes to sites of inflammation and the normal homing of lymphocytes
CC to secondary lymph organs.
XX
XX Sequence 386 AA;
SQ
Query Match 49.1%; Score 1008; DB 20; Length 386;
Best Local Similarity 52.3%; Pred. No. 6.9e-102;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;
QY 14 LLLAQTTCLLLFIISRP-----GPSSPAGGEDRVHVLVSSWRSGSFLGQ 59
Db 1 mllpkkmkllllflvsqmailalfhmyshnisslsmkagpermhvlvsswrsgsfvgq 60
QY 60 LFSQHDPDVFYLMPPAWHVVTTLSQGSAAATLHMAVRDLMRSLFCLCDMDVFDAYM-PQSRNL 118
Db 61 lfgqhpdvfylmepawhvwmtfkqstawlmlhmvdrdlrsvflcdmdvfdaymepgprq 120
QY 119 SAFTNWATSRALCSPACSAFPRTGTSKQDVCKTLCTROPFSLAREACRSYSHVVLKEVR 178
Db 121 sslqwnsralsapacdliipqdeiprahcrllcsqppfevvekakrsyshvvlkev 180
QY 179 FFNLQVLYPLSDPALNLRIVHLVRDPRAVLRSRERAGPILARDNGIVLGTN-GKWEAD 237


```
XX 08-DEC-1999 (first entry)
XX Mouse glycosyl sulfotransferase-3 protein sequence.
DE Glycosyl sulfotransferase; GSR-3; detection; diagnosis; leukocyte homing;
XX selectin binding interaction; inflammation; lymphocyte homing; mouse;
KW secondary lymph organ.
KW Mus sp.
XX WO9949018-A1.
OS 30-SEP-1999.
XX 26-FEB-1999; 99WO-US04316.
XX 20-MAR-1998; 98US-0045284.
PR 12-NOV-1998; 98US-0190911.
XX (REGC ) UNIV CALIFORNIA.
PA (SYNT ) SYNTEX USA INC.
XX Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
XX WPI; 1999-580442/49.
DR N-PSDB; AAZ20793.
XX Human and murine glycosyl sulfotransferase 3 and related
PT polynucleotides -
XX Claim 2; Fig 4; 59pp; English.
XX This sequence is the mouse glycosyl sulfotransferase-3 (GST-3) of
CC the invention. The nucleic acid sequences, probes and primers derived
CC from these, proteins and antibodies are useful in detecting homologues.
CC The sequences, antibodies and methods are useful in the diagnosis and
CC treatment of diseases associated with selectin binding interactions,
CC including conditions associated with or resulting from the homing of
CC leukocytes to sites of inflammation and the normal homing of lymphocytes
CC to secondary lymph organs.
XX Sequence 388 AA;
SQ
Query Match 46.7%; Score 958.5; DB 20; Length 388;
Best Local Similarity 55.6%; Pred. No. 1.9e-96;
Matches 193; Conservative 42; Mismatches 87; Indels 25; Gaps 6;
QY 42 VHVVLSSWRSGSFLGQLFSQHPDVFYLMPEAWHVWTTLSQGSAAATLHMAVRDLMRSIF 101
DB 42 vhlvlsswrsgrsfvgqlfgqhpdvfymlpeawhvwmtftsstawkhlmavrdllrsvf 101
QY 102 LCDMDVFDAYM-PQSRNLSAFFNWATSRALCSPACSAFPRGTISKQDVCKTLCTRPFS 160
DB 102 lcdmvsfdaymnpgrkgsslfqweqsralsapvcdfpaheisspkhckllcgqpf 161
QY 161 LAREACRSYSHVVLKEVRFENLQVLYPLLSDPALNLRIVHLVRDPRAVLRREAAGPILA 220
DB 162 mvekacsrshgfvvlkevrfslqalyplltdpslnlhvhlvrpavfrsrehttielm 221
QY 221 RDNQIVLGTNGKWEADPHLRILIRE-----VCRSHVRIAEEA-ATLKPPPPFLRGY 269
DB 222 v dshivlg-----qhletikeedqpyyamklicksqvdivkaiqt1--pealgqy 270
QY 270 RLVRFPEDLAREPLAEIRALYAFGLTLTPQLEAWIHNIHSGSIGKPIEAFTSSRNARN 329
DB 271 lflryedlvraplaqttrlykfvglfdflphlqtwvynvtrgkmgq--hafhtnarnaln 328
QY 330 VSQAWRHALPFTKILRVQEVCAQALQLLGYRPVYSADQQRDLTLDLV 376
DB 329 vsqawrwslypekvsqldacgeamdllgylqvrsgqeqgnlsldll 375
```

```
RESULT 8
AAY31656
ID AAY31656 standard; Protein; 483 AA.
XX
AC AAY31656;
XX 09-NOV-1999 (first entry)
XX Mouse N-acetylglucosamine-6-O-sulfotransferase.
DE
XX N-acetylglucosamine-6-O-sulfotransferase; mouse; GlyCAM-1;
KW L-selectin ligand.
XX
OS Mus musculus.
XX EP943688-A2.
XX 22-SEP-1999.
XX 04-MAR-1999; 99EP-0301530.
XX 24-JUN-1998; 98JP-0177844.
PR 05-MAR-1998; 98JP-0054007.
XX (SEKG ) SEIKAGAKU CORP.
XX Habuchi O, Kadomatsu K, Kannagi R, Muramatsu H;
PI Muramatsu T, Uchimura K;
XX WPI; 1999-520337/44.
DR N-PSDB; AAX87820.
XX New polypeptides of N-acetylglucosamine-6-O-sulfotransferase, useful
PT for synthesis of sugar chains, e.g. GlyCAM-1
XX Claim 1; Page 24-25; 41pp; English.
XX The present sequence represents mouse N-acetylglucosamine-6-O-
CC sulfotransferase, an enzyme capable of transferring a sulfate group
CC from a sulfate group donor to a hydroxyl group at the 6 position of
CC an N-acetylglucosamine residue located at the non-reducing end of
CC an oligosaccharide of formula GlcNAcbeta1-3Galbeta1-4GlcNAc, where
CC GlcNAc = N-acetylglucosamine residue, Gal = galactose residue,
CC beta 1-3 = beta 1-3 glycosidic linkage, and beta 1-4 = beta 1-4
CC glycosidic linkage. The enzyme is useful for the synthesis of
CC sugar chains such as GlyCAM-1, a ligand of L-selectin that is
CC involved in homing of lymphocytes and rolling of leukocytes
CC occurring at the early stage of inflammation. DNA encoding the
CC enzyme (see AAX87820) is expected to be used for the large-scale
CC production of N-acetylglucosamine-6-O-sulfotransferase, or
CC artificial synthesis of GlyCAM-1 using transformants which harbour
CC the DNA.
XX Sequence 483 AA;
SQ
Query Match 32.4%; Score 665; DB 20; Length 483;
Best Local Similarity 40.4%; Pred. No. 4.5e-64;
Matches 150; Conservative 64; Mismatches 121; Indels 36; Gaps 11;
QY 37 GGEDRVHVLVLSWRSGSFLGQLFSQHPDVFYLMPEAWHVWTTLSQGSAAATLHMAVRDL 96
DB 113 ggdkrqlvyvfttwsrgsffgelfngnpevf flyepvhwvqklypgdavslqgaardm 172
QY 97 MRSIFLCDMDVFDAYMPQ---SRNLS--AFPNWATSRALCSPACSAFPRGTISKQD--V 149
DB 173 lsalycrdlsvfqlyspagsggrnlttlgfgaatnkvvcssplocpayrkevglvdrv 232
QY 150 CKTLCTROPFSLAREACRSYSHVVLKEVRFENLQVLYPLLSDPALNLRIVHLVRDPRAVL 209
DB 233 ckk-cppqrlarfeecrkyrtvvikgvrfdvavlapllkdpaldlkvihlvrdprava 291
```

QY 210 RS-----REAAGPILARD-----NGIVLGT--NGKWEADPH-LRLIREV 246
 Db 292 ssrirsrhglireslqvrsrdprahrmpfleagghklgakkegmppadyhalgamevi 351
 QY 247 CRSHVRIAEAAATLKPPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPQLEAWIHN 306
 Db 352 cnsmaaktlqta-lqppdwlqghylvvryedlvdpvktlrrvydfvglivspemeqfaln 410
 QY 307 ITHGSG-IGKPIEAFHTSSRNARNVSOAWRHALPFTKILRQVEVCAGALQLLGYRPVISA 365
 Db 411 mtsgsgssskp---fvvsarnatqaanawrtaltfqikqveefcyqpmavlgervnsp 467
 QY 366 DQQRDLTLDLV 376
 Db 468 eevkdlsktll 478

RESULT 9
 AAB95367
 ID AAB95367 standard; Protein; 530 AA.
 XX AAB95367;
 AC
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:17679.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

Claim 8; SEQ ID 17679; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 CC represent oligonucleotides, all of which are used in the exemplification CC of the present invention.

SQ Sequence 530 AA;

Query Match 32.2%; Score 659.5; DB 22; Length 530;
 Best Local Similarity 39.1%; Pred. No. 2.1e-63;
 Matches 151; Conservative 68; Mismatches 122; Indels 45; Gaps 12;

QY 31 GPSSPAG-----GEDRVHVLVLSWSSGSSFLGQLFSQHPDVFLMEPAWHVWTTL 81
 Db 145 gvaappngtrgtgvgdkrqlvyvfttwtsgssffgelfnqnpevffiyepvhwvqkl 204
 QY 82 SQGSAATLHMAVYRDLMRSIFLCMDMDVFDAYMPQ---SRNLS--AFENWATSRALCSPAC 136
 Db 205 ypgdavsllggaardmlsalyrcdlsvfqlspagsggrnlttlgfgaatnkvvcssplc 264
 QY 137 SAFPRGTISKQD--VCKTLCITROPESLAREACRSYSHVVLKEVRFENLQVLYPLLSDPAL 194
 Db 265 payrkevvgldvdrvckk-cppqrlarfeecirkyrtlvikgvrfdvavlapllrdpal 323
 QY 195 NLRIVHLVRDPRAVLRS-----REAAGPILARD-----NGIVLGTNGKWV 234
 Db 324 dlkvihlvrdpravassrirsrhglireslqvrsrdprahrmpfleagghklgakkegv 383
 QY 235 --EADPH-LRLIREVCRSHVRIAEAAATLKPPFLRGYRLVRFEDLAREPLAEIRALYAF 291
 Db 384 ggpadyhalgamevicnsmaktlqta-lqppdwlqghylvvryedlvdpvktlrrvydf 442
 QY 292 TGLTLTPQLEAWIHNITHGSG-IGKPIEAFHTSSRNARNVSOAWRHALPFTKILRQVEVC 350
 Db 443 vglivspemeqfalnmtsgsgssskp---fvvsarnatqaanawrtaltfqikqveefc 499
 QY 351 AGALQLLGYRPVYSADQQRDLTLDLV 376
 Db 500 yqpmavlgervnspveevkdlsktll 525

RESULT 10
 AAY31657
 ID AAY31657 standard; Protein; 484 AA.
 XX
 AC AAY31657;
 XX
 DT 09-NOV-1999 (first entry)
 XX
 DE Human N-acetylglucosamine-6-O-sulfotransferase.
 XX
 KW N-acetylglucosamine-6-O-sulfotransferase; human; GlyCAM-1;
 KW L-selectin ligand.
 XX
 OS Homo sapiens.
 XX
 PN EP943688-A2.
 XX
 PD 22-SEP-1999.
 XX
 PF 04-MAR-1999; 99EP-0301530.
 XX
 PR 24-JUN-1998; 98JP-0177844.
 PR 05-MAR-1998; 98JP-0054007.
 XX
 PA (SEK) SEIKAGAKU CORP.
 XX
 PI Habuchi O, Kadamatsu K, Kannagi R, Muramatsu H;
 PI Muramatsu T, Uchimura K;
 XX
 DR WPI; 1999-520337/44.

XX PS Claim 4; Page 15-16; 21pp; English.

XX CC This sequence produces the protein KSGal6ST (Keratan sulphate

CC 6-sulphotransferase), which has a molecular weight of about 46700. The

CC cDNA sequence of this protein was obtained by radiolabelling the cDNA of

CC chick chondroitin 6-sulphotransferase (C6ST) and using this as a probe in

CC a random oligonucleotide-primed labelling method. Human foetal brain

CC cDNA was inserted into a phage Lambda gtl1 cloning vector whereby the

CC clones containing the KSGal6ST were obtained by hybridization using the

CC prepared probe. The positive clones were subcloned into a recombinant

CC expression vector and used to transform COS-7 cells, from which cells

CC expressing KSGal6ST can be selected. The phage cDNA inserts were

CC isolated and subcloned into a Bluescript plasmid. Deletion clones were

CC then prepared from which both strands were sequenced by the Sanger

CC method. The KSGal6ST of the invention transfers the sulphate from a

CC sulphate donor to galactose 6-OH groups in keratan sulphate, but does not

CC transfer sulphate to chondroitin, chondroitin sulphate A or C, dermatan

CC sulphate or CDSNS heparin.

XX PS Sequence 411 AA;

Query Match 25.0%; Score 513.5; DB 19; Length 411;

Best Local Similarity 33.5%; Pred. No. 1.6e-47;

Matches 139; Conservative 63; Mismatches 170; Indels 43; Gaps 12;

QY 7 SSKTVRVLLLA---QTTCLLLFIISR-----PGPSSPAGGEDRV-----H 43

Db 4 swkavilalasiaiaqytairtftaksfhtcpq-laeaglaericeesptfaynlrskth 62

QY 44 VLVLSWRSWSSGSSFLGQLFSQHPDVFLMEPAWHVWTTL-----SQGSA---AIIHMAVRD 95

Db 63 ililattrsgssfvqqlfnqhlvdvylfeplvhyvqntlprftqgkspadrvmलगasrd 122

QY 96 LMRSIFLCMDVDFDAYM---PQSRNLSAFFNWATSRALCSPACs-APPRGTISKQDVCK 151

Db 123 llrslvdcdlfleyikpppvnhdtfrfgrasrvlcsrpvcddppgpadlvleegdcv 182

QY 152 TLCTRQPFSLAREACRSYSHVVVKVEVFFNLQVLYPLLSDPALNLRIVHLVRDPRAVLRS 211

Db 183 rkcgllnltvaacacrershvaiktvrpvpevndlrvalvedprlnlkvglvrdprgilas 242

QY 212 REAAGPILARDNGIVLGTNGKWEADPHLRILIREVCRSHVRIAEAAATLKPPPLRGRYRL 271

Db 243 rsetfrdtyrlwrlwygtgrkypnld--vtqlttvcedfsnsvstgmlr-ppwlkgkym 299

QY 272 VRFDLAREPLAEIRALYAFTGLTLTPQLEAWIHNITHGS-GIGKPIEAFHTSSRNARNV 330

Db 300 vryedlarmpmkkteeiylgflpldshvarwigntrgdptlgk---hkygtvrnsaat 356

QY 331 SQAWRHALPFTKILRVQEVCAQALQLLGYRPVVSADQQRDLTLDLVLPGRGPDHFS 385

Db 357 aekwrfrlsydivafagnacqvlaqlgykiaaseeelknpsvslveerdfrpfs 411

RESULT 14

AAY72641

ID AAY72641 standard; Protein; 596 AA.

XX AC AAY72641;

XX DT 02 MAY-2001 (first entry)

XX DE Human glycosyl sulfotransferase-6 (GST-6) fragment.

XX KW Human; glycosyl sulfotransferase-6; GST-6; immunosuppressive;

KW therapy; selectin binding inhibitor; gene therapy; inflammation;

KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;

KW polyarthritis nodosa; polymyositis; systemic sclerosis; dermatitis;

KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis;

KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;

KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;

KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;

KW asthma; hypersensitivity; rheumatic fever; tissue rejection.

XX OS Homo sapiens.

XX PN WO200106015-A1.

XX PD 25-JAN-2001.

XX PF 19-JUL-2000; 2000WO-US19741.

XX PR 20-JUL-1999; 99US-0144694.

PR 13-JUL-2000; 2000US-0593828.

XX (REGC) UNIV CALIFORNIA.

XX PI Rosen SD, Lee JK, Hemmerich S;

DR WPI; 2001-138471/14.

DR N-PSDB; AAD02702, AAD02703.

XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for

PT diagnostic and therapeutic agent screening applications -

XX Claim 3; Fig 3; 128pp; English.

XX The present sequence is human glycosyl sulfotransferase-6 (GST-6)

CC fragment.

CC GST is a type 2 membrane protein useful for inhibiting a binding event

CC between a selectin and a selectin ligand, which comprises contacting the

CC selectin with a non-sulphated selectin ligand, GST and a small molecular

CC agent that inhibits the sulphation activity of GST. GST is also useful

CC in inhibiting a selectin mediated binding event. GST is useful in gene

CC therapy to treat disorders such as acute or chronic inflammation,

CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarthritis

CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,

CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's

CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious

CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,

CC dermatitis, myocarditis, regional enteritis, adult respiratory distress

CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,

CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection

CC during transplantation.

XX SQ Sequence 596 AA;

Query Match 5.8%; Score 119.5; DB 22; Length 596;

Best Local Similarity 21.6%; Pred. No. 0.00052;

Matches 86; Conservative 54; Mismatches 163; Indels 95; Gaps 16;

QY 28 SRPGSSPAGGEDRVHVLVLSSWRSWSSSFLGQLFSQHPDVFLMEPAWHVWTTLSQGSAA 87

Db 223 skkslssegghmdlpdvvitslpqsgaeilkqlfnnssdflyirvp-----ta 270

QY 88 TLHMAVRDLMRSIFL--CDMDVFDAYMPQSRNLSAFFNWATsRA-----LCsPP 134

Db 271 yidipeteleidsfvdacewkvsdi---rsghfrllrgwlqslvqdtklhlqnlhhep- 326

QY 135 ACSAFPRTISK-----QDVCKTLCTRQPFSLAREACRSYSHVVLKEVRFNQLVLYPLL 189

Db 327 -----nrgkiaqyfamkdkkrkfkrreslpeqrsqmkgafrdrdaeyiralrrhlvYyps 381

QY 190 SDPALNL-----RIVHLVRDPRAVLRsR-EAAGPILARDNGI--- 225

Db 382 arpvlslssgswtlklhffqevlgasmralyivrdprawiysmlynskpslyslknvpeh 441

QY 226 -----VLGTNGKW-----VEADPHLRILIREVCR-----SHVRIAE-AATLK-P 261

Db 442 laklfkieggkgkcnlnsgyafeyep---lrkelsksksnavslshlwlantaalrin 498

QY 262 PPFLRGRYRLVRFDLAREPLAEIRALYAFTGLTLTPQLEAWIHNITHGSIGKPIEAFH 321

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 18:18:06 ; Search time 315.85 Seconds
(without alignments)
3183.924 Million cell updates/sec

Title: US-09-593-828-4
Perfect score: 1173
Sequence: 1 atgtggctgccacggttc.....gtgggcacgcctgactga 1173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	1173	100.0	2044	22 AAD02699	Human glycosyl sul
2	1173	100.0	2170	22 AAD02698	Human glycosyl sul
3	1173	100.0	160552	22 AAD02697	Human glycosyl sul
4	1027.2	87.6	2988	21 AAC76156	Human ORFX ORF1711
5	924.6	78.8	1694	22 AAD02700	Human glycosyl sul
6	745.2	63.5	1989	22 AAD02696	Mouse glycosyl sul
7	400	34.1	2032	20 AA220792	Human glycosyl sul
8	377	32.1	2065	21 AA294211	Human transferrase
9	354	30.2	1926	20 AA220793	Mouse glycosyl sul
10	235.2	20.1	2156	19 AAV21200	Glycosaminoglycan
11	196.8	16.8	2354	18 AAT45037	Chick chondroitin

12	153.8	13.1	3029	22 AAH17922	Human cDNA sequenc
13	149.8	12.7	2409	20 AAX87821	Human N-acetylgluc
14	139	11.8	1458	19 AAV36418	Keratan sulphate 6
15	133	11.3	2150	20 AAX87820	Mouse N-acetylgluc
16	64.6	5.5	43280	18 AAT80413	Tylectone synthase
17	57.8	4.9	731	20 AAZ24563	Human lung tumor a
18	57.8	4.9	731	21 AAC65802	Human lung cancer-
19	57.2	4.9	30001	18 AAT61016	Total DNA sequenc
20	57.2	4.9	30001	20 AAX05110	S. aureofaciens DN
21	55	4.7	114955	20 AAX53491	Human adenosine A1
22	54.8	4.7	58857	21 AAA58471	Nucleotide sequenc
23	54.6	4.7	4689	21 AAZ87299	S. venezuelae macr
24	54.6	4.7	36778	21 AAZ87318	S. venezuelae pik
25	54.6	4.7	37948	21 AAZ87285	S. venezuelae pik
26	54.6	4.7	38506	21 AAA75633	Nucleotide sequenc
27	54.6	4.7	38506	21 AAZ56001	Recombinant cosmid
28	54	4.6	795	19 AAV55830	FIGA insert stabl
29	54	4.6	109519	22 AAS08693	Micromonospora DNA
30	53.8	4.6	1925	20 AAX90924	Epstein Barr Virus
31	53.4	4.6	53789	19 AAV21187	Amycolatopsis medi
32	53.2	4.5	1683	22 AAD03820	Pseudomonas fluore
33	52.8	4.5	1926	21 AAA50254	Epstein Barr virus
34	52.8	4.5	1926	22 AAF82902	EBV tethering prot
35	52.8	4.5	2580	21 AAA75454	Nucleotide sequenc
36	52.8	4.5	5452	20 AAX90923	Anti-sense strand
37	52.8	4.5	8705	20 AAZ23778	Vector pShuttle DN
38	52.8	4.5	9600	19 AAV21683	Vector plasmid pCM
39	52.8	4.5	10380	20 AAZ22248	Nucleotide sequenc
40	52.8	4.5	10596	14 AAQ51731	Plasmid pCisEBON f
41	52.8	4.5	10596	17 AAT40348	Plasmid pCisEBON f
42	52.8	4.5	10596	20 AAX15650	Nucleotide sequenc
43	52.8	4.5	16080	21 AAA59553	DNA clone pCEK C1,
44	52	4.4	1734	21 AAA40350	Swine pseudorabies
45	52	4.4	4897	11 AAQ03259	Pseudorabies virus

ALIGNMENTS

RESULT 1

AAD02699
ID AAD02699 standard; cDNA; 2044 BP.

XX AAD02699;

AC AAD02699;

DT 02-MAY-2001 (first entry)

XX Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #2.
Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation; systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes; polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis; glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis; Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia; demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis; myocarditis; adult respiratory distress syndrome; eczema; psoriasis; asthma; hypersensitivity; rheumatic fever; tissue rejection; chromosome 16q23.1; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 218..1390
/*tag= a
/product= "Human glycosyl sulfotransferase-4alpha (GST-4alpha)"
/note= "CDS is specifically claimed as SEQ ID NO: 4 in claim 6 (page no: 41) of the specification"

XX WO200106015-A1.

XX 25-JAN-2001.

PD


```
XX OS Homo sapiens.
XX FH
XX FT 5'UTR
XX FT
XX FT
XX FT CDS
XX FT
XX FT
XX FT /product= "Human glycosyl sulfotransferase-4alpha
XX FT (GST-4alpha)"
XX FT /note= "CDS is specifically claimed as SEQ ID NO: 4
XX FT in claim 6 (page no: 41) of the specification"
XX FT 1517..2134
XX FT /tag= c
XX PN WO200106015-A1.
XX PD
XX PD 25-JAN-2001.
XX PF
XX PF 19-JUL-2000; 2000WO-US19741.
XX PR
XX PR 20-JUL-1999; 99US-0144694.
XX PR 13-JUL-2000; 2000US-0593828.
XX PA
XX PA (REGC ) UNIV CALIFORNIA.
XX PI Rosen SD, Lee JK, Hemmerich S;
XX WI; 2001-138471/14.
XX P-PSDB; AAY72639.
XX PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX PT diagnostic and therapeutic agent screening applications -
XX PS Claim 6; Page 62; 128pp; English.
XX CC The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
XX CC alpha) cDNA. GST-4 gene is found on chromosome 16q23.1.
XX CC GST is a type 2 membrane protein useful for inhibiting a binding event
XX CC between a selectin and a selectin ligand, which comprises contacting the
XX CC selectin with a non-sulphated selectin ligand, GST and a small molecular
XX CC agent that inhibits the sulphation activity of GST. GST is also useful
XX CC in inhibiting a selectin mediated binding event. GST is useful in gene
XX CC therapy to treat disorders such as acute or chronic inflammation,
XX CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarthritis
XX CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
XX CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
XX CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
XX CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
XX CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
XX CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
XX CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
XX CC during transplantation.
XX SQ Sequence 2170 BP; 449 A; 682 C; 612 G; 427 T; 0 other;

Query Match 100.0%; Score 1173; DB 22; Length 2170;
Best local Similarity 100.0%; Pred. No. 2.5e-217;
Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgtggctgccacggttctccagcaagacagtgcagtgcctcctccctggcacagaccacc 60
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 61 tgcctcctgtcttctcatctcccgccagggccctcctatcccccagcgcgagagat 120
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 404 tgcctcctgtcttctcatctcccgccagggccctcctatcccccagcgcgagagat 463
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 121 cgtgtgcaogtgcgtgtcctcgtggcgctcgtgggctcatccttctgggccagctc 180
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 464 cgtgtgcaogtgcgtgtcctcgtggcgctcgtgggctcatccttcttgggccagctc 523
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
QY 181 ttcaagccagcaccacgctcttctacatgatggagcccgctggcatgtgtggaccacc 240
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 241 ctgtgcagggcagcgcggaacgctgcacatggccgtgcgcacctgatgcgtctatc 300
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 301 ttttgtcgacatggaagcgcggaacgctgcacatggccgtgcgcacctgatgcgtctatc 360
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 361 ttttcaactgggcaacgagccgcgcgtgtgtcgcgcgcgcgcgcgcgcgcgcgcgcgc 420
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 421 cgaggcaccatcagcaagcaggacgtatgcaagacacactgtgcacggcgacattcagc 480
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 481 ctggcccgaggagcctgcgcctcctacagccacgctggtgtcctcaaggaggtgcgtcttc 540
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 541 aaectgcaggtgctctacccgctgctcagcagcccccgcgcctcaacctgcgcacgtgcac 600
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 601 ctggtgcgcgaccccgccgctgctgcgcagccacgctgcgcagcccgccgatactggca 660
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 661 cgcgacaacggcactgctgtgggcaacaaacggaggtgggtggagggcgacccctacctg 720
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 721 cgcctgattcgcgaggtgtgcgcgacccacgctgcgcagcccgccgacactcaag 780
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 781 cgcgcaccttctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 840
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 841 ccgctggcagagatccgcgcactctacgccttccacggcctgacccctacgcacagctc 900
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 901 gaggcctggatcccaacacatcccccacgggtgcggggtgcggcaagccaatcgaggccttc 960
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 961 catacttgcctaggaatgcgcgaacgctctcctccagccctggcgcaacgctgtgccttc 1020
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1021 actaagatcctgcgcgtgcaggaggtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1080
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1081 cctgtgtactctgcggaccagcagcgtgacctcaccctggatctgtgtgcacagagc 1140
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1141 ccagaccacttcagctgggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1200
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1201 actaagatcctgcgcgtgcaggaggtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1260
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1261 cctgtgtactctgcggaccagcagcgtgacctcaccctggatctgtgtgcacagagc 1320
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1321 cgcctggcagagatccgcgcactctacgccttccacggcctgacctcaccgacacagctc 1380
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1381 gaggcctggatcccaacacatcccccacgggtgcggggtgcggcaagccaatcgaggccttc 1440
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1441 ccagaccacttcagctgggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1500
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1501 cctgtgtactctgcggaccagcagcgtgacctcaccctggatctgtgtgcacagagc 1560
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1561 cgcctggcagagatccgcgcactctacgccttccacggcctgacctcaccgacacagctc 1620
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1621 gaggcctggatcccaacacatcccccacgggtgcggggtgcggcaagccaatcgaggccttc 1680
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1681 cctgtgtactctgcggaccagcagcgtgacctcaccctggatctgtgtgcacagagc 1740
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1741 ccagaccacttcagctgggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1800
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1801 actaagatcctgcgcgtgcaggaggtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1860
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1861 cctgtgtactctgcggaccagcagcgtgacctcaccctggatctgtgtgcacagagc 1920
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1921 cgcctggcagagatccgcgcactctacgccttccacggcctgacctcaccgacacagctc 1980
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1981 gaggcctggatcccaacacatcccccacgggtgcggggtgcggcaagccaatcgaggccttc 2040
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2041 catacttgcctaggaatgcgcgaacgctctcctccagccctggcgcaacgctgtgccttc 2100
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2101 actaagatcctgcgcgtgcaggaggtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2160
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2161 cctgtgtactctgcggaccagcagcgtgacctcaccctggatctgtgtgcacagagc 2220
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2221 cgcctggcagagatccgcgcactctacgccttccacggcctgacctcaccgacacagctc 2280
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2281 gaggcctggatcccaacacatcccccacgggtgcggggtgcggcaagccaatcgaggccttc 2340
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2341 catacttgcctaggaatgcgcgaacgctctcctccagccctggcgcaacgctgtgccttc 2400
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2401 actaagatcctgcgcgtgcaggaggtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2460
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2461 cctgtgtactctgcggaccagcagcgtgacctcaccctggatctgtgtgcacagagc 2520
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

RESULT 3
AAD02697
ID AAD02697 standard; DNA; 160552 BP.
XX

Fri Feb 1 20:20:56 2002

AC AAD02697;
XX
DT 02-MAY-2001 (first entry)
XX
DE Human glycosyl sulfotransferase-4 (GST-4) genomic DNA.
XX
KW Human; glycosyl sulfotransferase-4; GST-4; immunosuppressive;
KW therapy; selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; anaemia;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; Hashimoto's
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW chromosome 16q23.1; ds.
XX
OS Homo sapiens.
XX
FH Location/Qualifiers
FT 32847..32922
FT exon /*tag= a
FT /number= 1
FT /label= 4a_504
FT 32923..35592
FT intron /*tag= b
FT /cons_splice= (5'site:NO, 3'site:YES)
FT 35593..35674
FT exon /*tag= c
FT /number= 2
FT /label= 4a_503
FT 35675..45093
FT intron /*tag= f
FT /cons_splice= (5'site:NO, 3'site:NO)
FT 46634..46700
FT exon /*tag= g
FT /number= 3
FT /label= 4a_502
FT 45186..46633
FT intron /*tag= h
FT /cons_splice= (5'site:YES, 3'site:NO)
FT 46701..47938
FT intron /*tag= i
FT /number= 4
FT /label= 4a_501
FT 46701..47938
FT intron /*tag= j
FT /cons_splice= (5'site:NO, 3'site:NO)
FT 47939..49746
FT exon /*tag= k
FT /number= 5
FT /note= "Includes 17 base pairs of 5'UTR, the ORF
FT and all of 3'UTR"
FT 47939..47955
FT 5'UTR /*tag= l
FT /note= "Portion of 5' untranslated region (5'UTR)"
FT 47956..49128
FT CDS /*tag= m
FT /product= "Human glycosyl transferase-4alpha
FT (GST-4alpha)"
FT 49129..49746
FT 3'UTR /*tag= n
FT /number= 1
FT 83257..83347
FT exon /*tag= o
FT /label= 4a_502
FT 83348..96412
FT intron /*tag= p
FT /cons_splice= (5'site:NO, 3'site:NO)
FT 96413..96484
FT exon /*tag= q
FT /label= 4a_501
FT 96485..98456
FT intron /*tag= r
FT /cons_splice= (5'site:NO, 3'site:NO)

FT exon 98457..99968
FT /*tag= q
FT /note= "Includes 17 base pairs of 5'UTR, the ORF
FT and all of 3'UTR"
FT 98457..98473
FT 5'UTR /*tag= r
FT /note= "Portion of 5' untranslated region (5'UTR)"
FT 98474..99661
FT CDS /*tag= s
FT /product= "Human glycosyl transferase-4beta
FT (GST-4beta)"
FT 99662..99968
FT 3'UTR /*tag= t
XX
PN WO200106015-A1.
XX
PD 25-JAN-2001.
XX
PD 19-JUL-2000; 2000WO-US19741.
XX
PR 20-JUL-1999; 99US-0144694.
PR 13-JUL-2000; 2000US-0593828.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Rosen SD, Lee JK, Hemmerich S;
XX
XX WPI; 2001-138471/14.
DR P-PSDB; AAY72639, AAY72640.
XX
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX diagnostic and therapeutic agent screening applications
XX
XX Example 1; Page 62-104; 128pp; English.
XX
XX The present sequence is human glycosyl sulfotransferase-4 (GST-4) genomic
XX DNA encoding GST-4alpha and GST-4beta. GST-4 gene is found on
XX chromosome 16q23.1.
XX GST is a type 2 membrane protein useful for inhibiting a binding event
XX between a selectin and a selectin ligand, which comprises contacting the
XX selectin with a non-sulphated selectin ligand, GST and a small molecular
XX agent that inhibits the sulphation activity of GST. GST is also useful
XX in inhibiting a selectin mediated binding event. GST is useful in gene
XX therapy to treat disorders such as acute or chronic inflammation,
XX systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
XX nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
XX glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
XX disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
XX anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
XX dermatitis, myocarditis, regional enteritis, adult respiratory distress
XX syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
XX bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
XX during transplantation.
XX
XX Sequence 160552 BP; 40281 A; 37573 C; 38015 G; 44564 T; 119 other;

Query Match 100.0%; Score 1173; DB 22; Length 160552;
Best Local Similarity 100.0%; Pred. No. 3.3e-217;
Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgtgggtgccacgggttctccagcaagacagtgacagtgctcctcctggcacagaccacc 60
Db 47956 atgtgggtgccacgggttctccagcaagacagtgacagtgctcctcctggcacagaccacc 48015
QY 61 tgcctcctgtcttctcatctcctccggccaggccctcatccccagcgcgaggat 120
Db 48016 tgcctcctgtcttctcatctcctccggccaggccctcatccccagcgcgaggat 48075
QY 121 cgtgtgcacgtgctgtgctgtcctcctcgtggcgtcggtcctcctcctcctcctcctc 180
Db 48076 cgtgtgcacgtgctgtgctgtcctcctcgtggcgtcggtcctcctcctcctcctcctc 48135

[illegible]

AC	AAC76156;	
XX		
DT	08-FEB-2001	(first entry)
XX		
DE	Human ORFX ORF1711	polynucleotide sequence SEQ ID NO:3421.
XX		
KW	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;	
KW	vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;	
KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;	
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;	
KW	hypotensive; dermatological; immunosuppressive; antiinflammatory;	
KW	antiviral; antibacterial; antifungal; antirheumatic; antithyroid;	
KW	antianaemic; gene therapy; cancer; proliferative disorder; hypertension;	
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;	
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;	
KW	cholesterol ester storage; systemic lupus erythematosus; infection;	
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;	
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;	
KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;	
KW	thrombosis; contraceptive; ss.	
XX		

Homo sapiens.

WO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US08621

31-MAR-1999; 99US-0127607

02-APR-1999; 99US-0127636.

05-APR-1999; 99US-0127728.

30-MAR-2000; 2000US-0540763.

(CURAGEN) CURAGEN CORP.

Shimkets RA, Leach M;

WPI; 2000-602362/57.

P-PSDB; AAB41947.

Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease.

Claim 5; Page 2597-2599; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human OREF open reading frames 1 to 3161. The OREF sequences have activities such as: cytostatic; hepatotropic; vulnery; antispasmodic; antiparkinsonian; nootropic; neuroprotective; osteopathic; anticonvulsant; antiarthritic; immunosuppressant; immunostimulant; cardiac; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatologic; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an OREF-associated disorder. The nucleic acids can be used to express OREF proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 2988 BP; 613 A; 916 C; 842 G; 617 T; 0 other;

Query Match

87.6%; Score 1027.2; DB 21; Length 2988;

RESULT 4
AAC76156
ID AAC76
XX

QY	118	gatcgtgtgcacgtgctggtgctgtcctcgtgagcgctcgggtcctcatcctctcttggggccag	177
Db	294	gagcgcatgcacgtgctggttctgtcttctcgtgagcgtctggtctctctcttcttcttggggccag	353
QY	178	ctcttcagccagcaccgcagctcttctacctgatgagagcccgctggcatgtgtggacc	237
Db	354	ctttttggcagcaccacagatgtttctacctgatgagagcccgctggcacgtgtggtatg	413
QY	238	acctgtcgcagggcagcgggcacgcctgcacatggccgtgctgcgcgacctgatgcgcctct	297
Db	414	accttaagcagagcacgcgcctggatgctgcacatggctgtgctgagagcccgctgatacgggcc	473
QY	298	atctttttgtcgcacatggacgtgtttgatgcctacatgcacagag--ccgaaacctg	354
Db	474	gtctttgtcgcacatgacgtctcttgatgcctacatgcctacatggaaacctgtcccgagagacg	533
QY	355	tccgccttttcaactgggcaacgagccgcgcgtgtgctgcgcgcgcgcctgcagcgcc	414
Db	534	tccagcctctttcaagtggagaaacgcgcgcgtgtgttctgcacctgcctgtgacatc	593
QY	415	tttcccgagccaccatcagcaagcaggaacctgtgcaagacactgtgcacgcggcag-cc	473
Db	594	-atcccaacagatgaatcatcccccggtctactgcaggtcctgtgcagtcacacagccc	652
QY	474	attcagcctggccgggagggcctgcgcgtcctacagccactgtgtgctcaaggaggtgcg	533
Db	653	tttgaagtgttggaagaaacctgcgcgtcctacagccacgtgtgtcctcaaggaggtgcg	712
QY	534	cttcttcaacctgcaggtgctctacccgctgctcagcagcccccgcgtcctcaacctgcgcac	593
Db	713	cttcttcaacctgcagtcctctacccgctgctgaaagacccctccctcaacctgcacat	772
QY	594	cgtgcacctggtgcgcgacccgcggccgctgctgcgcgtcccgaggcggcgcccgat	653
Db	773	cgtgcacctggtccgggaacccccggccgctgttccgttcccgagaaacgacaaagggaga	832
QY	654	actggcacgcgacaacggcatcgtgctgggcacaaaggcaagtgggtggagggccgaccc	713
Db	833	tctcatgattgacagtcgcattgtgtatggggcagcatgagcagaaactcaagaaggaggga	892
QY	714	tcacctggcgcctgattcgcgaggtgtgcgcgagccactgctgcgcagcccgccac	773
Db	893	ccaacctactatgtgatgcaggtcattctgcctacatgcctacacgagcagtcacagacat	952
QY	774	actcaagccgcaccccttctcgcgggcgcctaccgcctgtgtgcgttgcgcttgcagaccctggc	833
Db	953	ccagtcttgcctcaaggccctgcaggaaacgctacacgtctgtgtgcgtatgagcacctggc	1012
QY	834	gcgggagccgctggcagagatccgcgcactctacgccttccaccggcctgacctcagcc	893
Db	1013	tcgagccctgtggcccagacttcccgaaatgatgaattcgtgggattgggaattcttgc	1072
QY	894	acagctcagggcctggatcccaacaatcacccacgggtcgggggatcggaagcacaatcga	953
Db	1073	ccatcttcagaaacctgggtgcataaacatcacccgaggaaggccatgggtga-----cca	1126
QY	954	ggccttccatacttctgttaggaatgcgcgcaacgtctccacggcctggcgccacgcgtt	1013
Db	1127	cgtttccacacaaaatgccagggtgccccttaattgtctcccaggccttggcgctggtcttt	1186
QY	1014	gcccttcaataagatcctgcgcgtgcaggaggtgtgtgcgcggcgcgctgcagctgctggg	1073

Db 1086 gtgtgcaagaagtgcgcgccacagcgccgtggttcgagggaggagtgccgcgaagtac 1145
QY 508 agccacgtgtgtctcaaggaggtgcgcttcttcaacctgcaggtgctctacccgtgctc 567
Db 1146 cgcacactagtataaagggtgtgcgcttctcgacgtggcgttctggcgccactgtg 1205
QY 568 agcgaccgcggtcacaacctgcgcacatcgtgcacactggtgcgcgacccggggccgtgctg 627
Db 1206 cgagaccgcgccctggacctcaaggtcatccacttggctgactcccgcgcggtggcg 1265
QY 628 cgctcccgaggcgcgcgcccgccgatactggcacgcgacaac----- 669
Db 1266 agttcacggatccgctcgcgcacagcgccctcatccgtgagagcctacaggtggtgcgcagc 1325
QY 670 -----ggcatcgtgctgggcacc 687
Db 1326 cgagaccgcgagctcaccgcatgcccttcttggaggccgcgggcccacagcttggcgcc 1385
QY 688 aacggcaagtgggtgaggcgacccctcacctgcgcctgattcgcgaggtgtgcgcgacg 747
Db 1386 aagaaggaggcgtggcgcccgccgcagactaccacgctctggtggtatggggtatc 1445
QY 748 cccgtgcgcacgcgcgagc-----cgccacactcaagccgcaccccttctcgcgcgc 801
Db 1446 tgcaatagtatggctaagacgctgcagacagccctgcagcccccctgactggctgcagggc 1505
QY 802 cgcctaccgcctggtgcgtcttcgaggacctggcgcgaggccgctggcagagatccgcgca 861
Db 1506 cactacctggtggtgcggtacgaggacctggtgggagaccccgctcaagacactacggaga 1565
QY 862 ctctacgccttcacgcgccctgacccctcaogccacagctcgaggccctggatccacaacatc 921
Db 1566 gtgtacgattttgtggactgttggtagcccgcaaatggagcagtttgcctgaacatg 1625
QY 922 accacagggtcgggagatcggcaagccaatcgaggcccttccatacttctgtaggaatgcg 981
Db 1626 accagtggctcg-----ggctcctcctccaagcccttctggtatctgcacgcaatgcc 1679
QY 982 cgcaacgtctccaggccctggcgccacgcggttgcccttcaactaagatcctgcgctgcag 1041
Db 1680 acgcaggcgccaatgcctggcgagccgcccctcaccttccagcagatacaaacaggtggag 1739
QY 1042 gaggtgtgcgcggcgctgcagctgctgggctaccggcctgtgtactctgcggaccag 1101
Db 1740 gagtttgcctaccagcccatggcgtcctggctatgagcgggtcaacagccctgaggag 1799
QY 1102 cagcgtgacctcaccctggatctggtgctg 1131
Db 1800 gtcaagacctcagcaagaccctgcttcgg 1829

RESULT 14
AAV36418
ID AAV36418 standard; cDNA; 1458 BP.
XX
AC AAV36418;
XX
DT 14-SEP-1998 (first entry)
XX
DE Keratan sulphate 6-sulphotransferase.
XX
KW KSGal6ST; keratan sulphate 6-sulphotransferase; hybridization;
KW chick chondroitin 6-sulphotransferase; C6ST; phage Lambda;
KW expression vector; COS-7 cells; Bluescript plasmid; galactose;
KW keratan sulphate; chondroitin; ss.
XX
OS Homo sapiens.
XX
XX
Key Location/Qualifiers
CDS 1..1236
/*tag= a
/product= "Keratan sulphate 6-sulphotransferase"

PN EP845533-A2.
XX
PD 03-JUN-1998.
XX
PF 27-NOV-1997; 97EP-0309564.
XX
PR 29-NOV-1996; 96JP-0320535.
XX
PA (SEK) SEIKAGAKU KOGYO CO LTD.
XX
PI Fukuta M, Habuchi O;
XX
DR WPI; 1998-288750/26.
XX
P-PSDB; AAW61100.
PT Keratan sulphate 6-sulpho-transferase polypeptide - transfers
PT sulphate from sulphate donor to galactose 6-hydroxy group etc.
XX
PS Claim 9; Page 15-16; 21pp; English.
XX
CC This nucleic acid sequence is a cDNA coding for human KSGal6ST (Keratan
CC sulphate 6-sulphotransferase). The sequence was obtained by
CC radiolabelling the cDNA of chick chondroitin 6-sulphotransferase (C6ST)
CC and using this as a probe in a random oligonucleotide-primed labelling
CC method. Human foetal brain cDNA was inserted into a phage Lambda gt11
CC cloning vector whereby the clones containing the KSGal6ST were obtained
CC by hybridization using the prepared probe. The positive clones were
CC subcloned into a recombinant expression vector and used to transform
CC COS-7 cells, from which cells expressing KSGal6ST can be selected. The
CC phage cDNA inserts were isolated and subcloned into a Bluescript plasmid.
CC Deletion clones were then prepared from which both strands were sequenced
CC by the Sanger method. The KSGal6ST of the invention transfers the
CC sulphate from a sulphate donor to galactose 6-OH groups in keratan
CC sulphate, but does not transfer sulphate to chondroitin, chondroitin
CC sulphate A or C, dermatan sulphate or CDSNS heparin.
XX
SQ Sequence 1458 BP; 254 A; 526 C; 421 G; 257 T; 0 other;

Query Match 11.8%; Score 139; DB 19; Length 1458;
Best Local Similarity 49.8%; Pred. No. 2.4e-18;
Matches 531; Conservative 0; Mismatches 485; Indels 51; Gaps 5;
QY 127 cacgtgctggtgctcctcgtggcgctcggtgcctcctcttcttggccagctcttcagc 186
Db 184 cacatcctcatcctggccaccacgcgcgcgcctcctccttcttggccagctcttcaac 243
QY 187 cagcaccgcgcgcctcttctacatgatggagcccgctggcatgtgtggaccaccct---- 242
Db 244 cagcacctggacgtcttctacactgtttgagccctctacacgctccagacacgcctc 303
QY 243 -----gtcgcaggggcagcgcgccgcaacgctgcacatggccgtgcgc 282
Db 304 ccccgcttaccaccagggcaagagcccgccgaccgcggtgcgtgtaggcgcgcgcgc 363
QY 283 gacctgatgcgctctatctttttgtgcgacatggacgtgtttgatgcctacat----- 335
Db 364 gacctcctgcggagcctctacgactgcgacctctacttctctggagaactacatcaagccg 423
QY 336 --gccacagagccgaaacctgtccgccttttcaactgggcaacgagccgcgctgtgc 393
Db 424 ccgcccgtcaaccacacacacagcagcaggtcttccgcggggccagccgggtcctctgc 483
QY 394 tcgcgcgcgcgcctgcagcgccttcccccgaggccacc---atcagcaagcaggacgtatgc 450
Db 484 tcccgccctgtgtgcgacctccggggccagccgacctgtcctggaggaggggactgt 543
QY 451 aagacactgtgcacgcggcagccattcagcctggccgggaggcctgcgcctcctacagc 510
Db 544 gtgcgcaagtgcgggtactactcaacctgacccgtgcgcggcggcgtgcgcgagcgcagc 603
QY 511 cacgtggtgctcaaggagggtgcgcttcttccaaacctgcaggtgctctaccccgctgc 570

Db 604 cacgtggccatcaagaacggtgcggtgcccggaggtgaacgacctgcgcgccctggtggaa 663

QY 571 gaccccgcgctcaacctgcgcctgctgcacctggtgcgcgaccccgccggtgctgcgc 630

Db 664 gaccgcgattaaacctcaaggtccatcagctggtccgagacccccgcgcattctgct 723

QY 631 tccgggagcgcgccgccccgatactggcacgcgacaacggcagctgctgggcaccaac 690

Db 724 tcgcgcagcgagaccttccgcgacacg-----taccggctcgtggcgctctggtac 774

QY 691 ggcaagtgggtggagggccgacctcaccctgcgcctgattcgcgaggtgtgcccagccac 750

Db 775 ggcacccgggaggaacacctacaacctggacgtgaacgagctgaccacggtgtgcgaggac 834

QY 751 gtgcgcacgcgcgagggccgcacactcaagccgaccccttccctgcgcgcgcgctacgcg 810

Db 835 ttctccaaactccgtgtccaccggcctcatgcggcccccggtggtccaaaggccaagtacatg 894

QY 811 ctggtgcgcttcgaggaacctggcggggagccgctggcagagatcccgcgccactctacgcc 870

Db 895 ttggtgcgctacgaggaacctgctcggaacctatgaagaagaccggaggagatctacggg 954

QY 871 ttacccggcctgacctcaacgcacagctcgagggcctggatccacaacatcacccacggg 930

Db 955 ttctgggcatacccgctggacagccacgtggcccccggtggccgctggatccagaacaacac-----g 1008

QY 931 tcggggatcggaagccaatcgagggccttccatacttctgtctaggaatgcgcgcaacgctc 990

Db 1009 cgggggacccccacctgggcaagcacaaatacggcacgtgcgaacctcgcgggccaag 1068

QY 991 tccaggcctggcgccacgcgttgcccttcaactaagatcctgcgcgtgcaggaggtgtgc 1050

Db 1069 gccgagaagtggcgcttcgcctctctctacgacatcgtggcctttgcccagaacgcctgc 1128

QY 1051 gccggcgcgctgcagctgctgggctaccggcctgtgtactctgcggaccagcagcgtgac 1110

Db 1129 cagcaggtgctggcccagctgggtctacaagatgcgcgcctcgaggaggagctgaagaac 1188

QY 1111 ctacacctggatctggtgctgccacgagggcccgaccagaccacttcaagt 1157

Db 1189 cctcggtcagcctggtggaggagcgggacttccgcccccttctctgtg 1235

RESULT 15

AAx87820

ID AAX87820 standard; cDNA; 2150 BP.

AC AAX87820;

XX 09-NOV-1999 (first entry)

DT Mouse N-acetylglucosamine-6-O-sulfotransferase cDNA.

XX N-acetylglucosamine-6-O-sulfotransferase; mouse; GlyCAM-1;

KW L-selectin ligand; ss.

XX Mus musculus.

XX Key Location/Qualifiers

FH 470..1921

FT /*tag= a

FT

XX EP943688-A2.

PN 22-SEP-1999.

XX 04-MAR-1999; 99EP-0301530.

PF 24-JUN-1998; 98JP-0177844.

XX 05-MAR-1998; 98JP-0054007.

PR (SEK) SEIKAGAKU CORP.

XX

PI Habuchi O, Kadomatsu K, Kannagi R, Muramatsu H;

PI Muramatsu T, Uchimura K;

XX

DR WPI; 1999-520337/44.

DR P-PSDB; AAY31656.

XX

PT New polypeptides of N-acetylglucosamine-6-O-sulfotransferase, useful

PT for synthesis of sugar chains, e.g. GlyCAM-1

XX

XX Claim 5; Page 21-23; 41pp; English.

PS

XX This is the nucleotide sequence of an isolated cDNA coding for

CC mouse N-acetylglucosamine-6-O-sulfotransferase (see AAY31656), an

CC enzyme capable of transferring a sulfate group from a sulfate group

CC donor to a hydroxyl group at the 6 position of an N-acetylglucosamine

CC residue located at the non-reducing end of an oligosaccharide

CC represented by the formula GlcNAcbetal-3Galbetal-4GlcNAc, where

CC GlcNAc = N-acetylglucosamine residue, Gal = galactose residue,

CC beta 1-3 = beta 1-3 glycosidic linkage, and beta 1-4 = beta 1-4

CC glycosidic linkage. The cDNA was isolated from a mouse embryo

CC cDNA library by PCR amplification. The enzyme is useful for the

CC synthesis of sugar chains such as GlyCAM-1, a ligand of L-selectin

CC that is involved in homing of lymphocytes and rolling of leukocytes

CC occurring at the early stage of inflammation. DNA encoding the

CC enzyme is expected to be used for the large-scale production of

CC N-acetylglucosamine-6-O-sulfotransferase, or artificial synthesis

CC of GlyCAM-1 using transformants which harbour the DNA.

XX

SQ Sequence 2150 BP; 386 A; 695 C; 679 G; 390 T; 0 other;

Query Match 11.3%; Score 133; DB 20; Length 2150;

Best Local Similarity 54.3%; Pred. No. 3.5e-17;

Matches 324; Conservative 0; Mismatches 255; Indels 18; Gaps 2;

QY 109 ggcggcgagagatcgtgtgcacgtgctgtgtcctcgtggcgctcggtcctccttc 168

Db 806 ggaggggacaagcggcagttggtgtatgtgttcaccacgtggcgctcggtcgtccttc 865

QY 169 ttggggccagctcttcagccagcaccccgagcgtcttctacctgatggagccgcgtggcat 228

Db 866 ttcggtgagctcttcaaccagaacctgaggtgtgtttctcctatgagcctgtgtggcac 925

QY 229 gtgtggaccacctgtcgcagggcagcggcggcaacgctgcacatggcgtgcgcgacctg 288

Db 926 gtgtggcaaaaactgtaccccggggagcgcgtttcctgcaggggcagcgcgggagacatg 985

QY 289 atgcgctctatctttttgtgcacatggacgtgtttgatgcctacatgcc----- 338

Db 986 ctgagcgctctctaccgctgcgatctttcgttttccagctgtatagccccgcaggcagt 1045

QY 339 -----acagagccgaaacctgtccgcctttttcaactgggcaacgagccgcgctgtgc 393

Db 1046 ggggggcgaacctcaccactctggcatcttttgggcagccactaacaagtggtatgc 1105

QY 394 tcgcccgcgcctgcagcgccttttccccgaggcaccatcagcaagcaggacgtatgcaag 453

Db 1106 tctcgcactctgtcctgcctaacgcctggcacgcttcgcaggttcgaggtggtgacgcgcgtg 1165

QY 454 ---acactgtgcacgggcagccattcagcctggcccgggagcctgccgctcctacagc 510

Db 1166 tgcaaaaagtgcccacctcaacgcctggcacgcttcgcaggttcgaggtggtcgaagtaccgc 1225

QY 511 cagtggtgctcaaggaggtgcgcttcttcaacctgcaggtgtctctaccgcgtcagc 570

Db 1226 acggtgtatcaaggcgctgcgggtctctcagatgtggtgtgttgccgctgtctaaa 1285

QY 571 gaccccgctcaacctgcgcacgtgcacacctggtgcgcgaccgcggcgctgtgcgc 630

Db 1286 gatccagccttggaacctcaaggttcacacctagtagctgacgtgctgtgtgtgccagc 1345

QY 631 tccccggagcgcgccccgatactggcacgcgacaaacggcatcgtgtgtgggcacc 687

Db 1346 tccgcacccgctcgcgtcacggccctccatccgggaaagcctacaggtggtgcgaagc 1402

Search completed: January 31, 2002, 18:19:26
Job time: 9120 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 16:44:48 ; Search time 3220.32 Seconds
(without alignments)
3914.145 Million cell updates/sec

Title: US-09-593-828-4
Perfect score: 1173
Sequence: 1 atgtggctgccacggttctc.....gctgggcacgtgcgtgactga 1173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estfun:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estom:*
 - 5: em_estpl:*
 - 6: em_estba:*
 - 7: em_estro:*
 - 8: em_estov:*
 - 9: em_htc:*
 - 10: gb_est1:*
 - 11: gb_est2:*
 - 12: gb_htc:*
 - 13: gb_gss:*
 - 14: em_gss_fun:*
 - 15: em_gss_hum:*
 - 16: em_gss_inv:*
 - 17: em_gss_pln:*
 - 18: em_gss_pro:*
 - 19: em_gss_rod:*
 - 20: em_gss_vrt:*
 - 21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	657.4	56.0	695	11 BE857538	BE857538 7g01a08.x
C 2	631	53.8	735	11 BE858652	BE858652 7g01a09.x
C 3	525	44.8	525	11 BF197521	BF197521 7o84a08.x
C 4	487.6	41.6	620	10 AI824100	AI824100 wj46c01.x
C 5	420.8	35.9	436	10 AW081348	AW081348 xc41b06.x
C 6	352.4	30.0	1923	12 AK009113	AK009113 Mus muscu
C 7	346.4	29.5	1067	13 CNS03KN7	AL248380 Tetraodon
C 8	290.6	24.8	568	10 AI115260	AI115260 ui43c07.y
C 9	278	23.7	849	13 CNS04QFN	AL302540 Tetraodon
C 10	271.2	23.1	517	11 BF042384	BF042384 BP250022A
C 11	240.2	20.5	433	13 AZ405100	AZ405100 IM0173108
C 12	224.4	19.1	695	11 BG964671	BG964671 602831875

13	224.4	19.1	783	11	BG963298	BG963298 602827716
14	218.8	18.7	852	11	BG966340	BG966340 602832826
15	159.4	13.6	965	11	BF579746	BF579746 602095056
16	148.6	12.7	322	10	AI156825	AI156825 ui44c08.y
17	134.6	11.5	2070	12	AK011202	AK011202 Mus muscu
18	134.4	11.5	793	11	BI102274	BI102274 602885587
19	130.6	11.1	955	11	BG107354	BG107354 602290740
20	129.2	11.0	634	10	AU180328	AU180328 AU180328
21	116.2	9.9	662	11	BF344303	BF344303 602017370
22	111.2	9.5	1070	11	BF163765	BF163765 601769868
23	109.2	9.3	540	10	AW412223	AW412223 uq47g02.y
24	103	8.8	103	10	AI282873	AI282873 qt87e06.x
25	103	8.8	297	10	AA261202	AA261202 va49a06.r
26	103	8.8	500	10	AI529474	AI529474 va49a06.y
C 27	99.8	8.5	135	10	AI824198	AI824198 wj36a01.x
28	97.4	8.3	494	10	AI088880	AI088880 gal7a07.x
29	97.4	8.3	539	11	BF593996	BF593996 nac20a08.
30	97.4	8.3	604	11	BE857485	BE857485 7f98g02.x
31	97.4	8.3	722	10	AW027448	AW027448 wv73h09.x
32	96.6	8.2	714	11	BI115837	BI115837 602866294
33	96.4	8.2	675	10	AI939595	AI939595 tf88ell.x
34	94.2	8.0	468	13	AQ939450	AQ939450 NR1-153R
C 35	84.2	7.2	401	13	AQ017697	AQ017697 CIT-HSP-2
36	82.2	7.0	607	11	BG971346	BG971346 602839896
37	81.2	6.9	284	10	AA243229	AA243229 zr26d04.r
38	78.6	6.7	516	10	BE208539	BE208539 ba08e09.y
39	78.6	6.7	965	11	BE794921	BE794921 601589671
C 40	78.6	6.7	1094	10	AL576297	AL576297 AL576297
41	77.2	6.6	503	11	BF323161	BF323161 maa37c02.
42	76	6.5	550	10	BE754360	BE754360 207854 MA
43	74.6	6.4	542	10	AL120609	AL120609 DKF7p761G
C 44	72.4	6.2	932	13	CNS0072Q	AL066742 Drosophil
45	70.8	6.0	581	11	BI346952	BI346952 376327 MA

ALIGNMENTS

RESULT 1
BE857538/c
LOCUS BE857538 695 bp mRNA EST 29-SEP-2000
DEFINITION 7g01a08.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3305174 3'
similar to TR:Q9UBY3 Q9UBY3 N-ACETYLGLUCOSAMINE
6-O-SULFOTRANSFERASE. ; mRNA sequence.
ACCESSION BE857538
VERSION BE857538.1 GI:10371664
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 695)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrta R. Rosenfeld M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 431.
Location/Qualifiers 1..695
/organism="Homo sapiens"

QY	544	ctgcagg-tgcttaccgcgtgctcagcgaccgccgcgctcaacctgcacctgctgcacct	602
Db	256	CTGCAGGTTGCTTACCCGCTGCTCAGCGACCCCGCGCTCAACCTGCGCATGCTGCACCT	197
QY	603	ggtgcgcgaccgcggccgtgctgcgtccgcctcccgaggcgggcgccgatactggcacg	662
Db	196	GGTGCGCGACCGCGCGCGCTGCTGCGTCCCGGAGCGGGCGCGCCGATACTGGCACG	137
QY	663	cgacaacggcatgctgctgggcaccacgaagtggtggagcgagccctcacctgog	722
Db	136	CGACAACGGCATGCTGCTGGGCACCAACGCAAGTGGTGGAGGCGACCCICACCTGCG	77
QY	723	cctgattcgcgaggtgtgcgcgagccacgtgcgcacgcccgcagccgcgcacactcaagcc	782
Db	76	CCTGATTGCGAGGTGTCCCGCAGCCACGTGCGCATCGCGGAGCGCGCCACACTCAAGCC	17
QY	783	gccaccttccctgcgc	798
Db	16	GCCACCTTCCTGCGC	1
RESULT	6		
AK009113			
LOCUS			
DEFINITION	AK009113	1923 bp mRNA HTC	05-JUL-2001
ACCESSION		Mus musculus adult male tongue cDNA, RIKEN full-length enriched	
VERSION	AK009113	library, clone:2310003G18, full insert sequence.	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			

Fri Feb 1 20:21:00 2002

Db 339 ACCAGCAGCACAGCCTGGAAGCTGCACATGGCTGTGCGGGATCTTCTCGGTTCCGTCCTTC 398
QY 304 ttgtgcacatggacgtgtttgatgcctacatg---ccacagagccgaaacctgtccgcc 360
Db 399 CTGTGTGACATGAGCGTCTTTGATGCCTACAIGAACCAGGCCCGGAAACAGTCCAGC 458
QY 361 tttttcaactgggcaacgagccgcgcgtgtgtctgcgcgcgcgcctgcagcgccctttccc 420
Db 459 CTCTTCCAGTGGGAGCAAAAGCGGGCCCTGTGCTCAGCGGCTGTGTGTGACTTCTTCCCT 518
QY 421 cgaggcaccatcagcaagcagacgtatgcaagacacactgtgcacgcggcagccattcagc 480
Db 519 GCCACGAGATCAGCTCACCCACACTGCAAGCTGCTCTCGGTCAGCAGCCCTTTGAT 578
QY 481 ctggccgggaggccctgcgcctctacagccacagtgtgtctcaaggaggtgcgctttctc 540
Db 579 ATGGTGGAGAAGGCTTCCGCTCTCACGGCTTCGTGCTACTCAAGGAGGTGCGTTTCTC 638
QY 541 aacctgaggtgtctacccgctgtcagcagcccgcgctcaacctgcgcacatcgtgcac 600
Db 639 AGCCTGAGGCCCTCTATCCACTACTACAGGACCCCTTCCCTCAACCTGCACGTCGTGCAC 698
QY 601 ctggtgcgcagcccgggcgctgtgcgtctccgggagggcgcgccgatacttgcca 660
Db 699 CTGGTCCGAGACCCCGGGCCGTGTTCCGATCCGGGAGCACACCACCATAGAACTCATG 758
QY 661 cgcgacaacgcatgtctgtggcaccacaggaagtgggtggaggccagccctcacctg 720
Db 759 GTTGACAGTCAATATGTCTAGGGCAGCAATTTGGAACGATCAAGGAGGAGACACGCC 818
QY 721 cgcctgattcgcaggtgtgccgcagccacagtgcgcacgtgcgcagggcgccgcaactcaag 780
Db 819 TATTATGCCATGAGATCATCTGCAAAAGCCAGGTGGACATAGTCAAGGCCATCCAAACC 878
QY 781 cggccacaccttctcggcgccgctaccgcctgtgtcgcttcgaggaacctggcggggag 840
Db 879 CTCCCTGAAGCTTGCAGCAGCGCTACCTGTTCCCTAGGTATGAGGACCTGGTTCGGGCA 938
QY 841 ccgctggcagagatccgcgcactctacgccttcaccggcctgacctcacgcccacagctc 900
Db 939 CCCTGGCCACGACGACCATATATAAATTTGTGGGTTGGAATTTTGGCCCCACCTC 998
QY 901 gaggcctggatccacaacatcaccacgggtgcggcgccgcgcgtgcagcgaatcgaggccttc 960
Db 999 CAAACATGGGTTTACAATGTACCCCGCGCAAGGGCATGGGTACG-----ATGCTTTC 1052
QY 961 catacttgctaggaatgcgcgaactctcccaggcctgcgcgaacgcgttgcccttc 1020
Db 1053 CATACTAACGCCAGGAACGCCCTCAACGCTCTCTCAGGCGTGGCGTGGTCTTACCTTAC 1112
QY 1021 actaatcctgcgcgtgcagaggtgtgcgcggcgccgcgcgtgcagcgtgcggctaccgg 1080
Db 1113 GAAAGGTTTCCAGCTTCAAGATGCCTCGGTCGAGGCTATGGATTGCTGGGATACCTC 1172
QY 1081 cctgtactctgcggaccagcagcgtgacctcacccctggatctggtg 1128
Db 1173 CAGGTCAGATCTCAACAGAACAGGCAACCTGTCCCTGGATCTTCTG 1220

RESULT 7
CNS03KN7/c 1067 bp DNA GSS 17-MAY-2000
LOCUS
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 033J20 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL248380
VERSION AL248380.1 GI:7969392
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;

Tetraodontidae; Tetraodon.
1 (bases 1 to 1067)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 1067)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 1067)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source
Location/Qualifiers
1..1067
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="033J20"
/clone_lib="G"
/note="Genoscope sequence ID : C0BG033DE10LPI-end : T7"
BASE COUNT 241 a 306 c 323 g 197 t
ORIGIN

Query Match 29.5%; Score 346.4; DB 13; Length 1067;
Best Local Similarity 62.1%; Pred. No. 3e-55;
Matches 564; Conservative 0; Mismatches 341; Indels 3; Gaps 1;
QY 17 tctccagcaagacagtgaacgtgtctctctctctggcacagaccactgctcctctctctca 76
Db 906 TCGGCACCATGATTTTGTGTTGGTGACCTCCAGGGAGCCACGGTGATGCTGTGAGGGCT 847
QY 77 tcattctccggccaggccctcatcctctctctctggcacagaccactgctcctctctctca 136
Db 846 GGTACTTCCAGCTCAGCCCCCTGCTCCTCGCTCCTCCTCGGGGAAAGTTACAGTCTCTCC 787
QY 137 tgcgtcctcgtggcgctcgggtcctctctctctctggccagctcttcagccagcaccccg 196
Db 786 TGCTGTCGTCCTGGCGCTCGGGCTCGTCCCTCCTGGGTGAGGTGTTTCAGTCAGCACCCGT 727
QY 197 acgtcttctacatgagcccgcggtgacatgtgtggaccacccctgtcgaggggcagcg 256
Db 726 CGGTCTTCTATCTGATGAGCCCGGGTGGCACGCTGTGGAGCCAGCTCCCAAGGTGAACG 667
QY 257 cggcaacgctgcacatggccgtgcgcgacactgatgcgctctatctttttgtggacatgg 316
Db 666 TTCGGATGCTCCGGATGCTGTCAAGGATATATTTCCGAGCATATTCAGTSCGACTTCT 607
QY 317 acgtgtttgatgcctacatgccacagagccgaaacctgtccgcctttttcaactgggcaa 376
Db 606 CAGGAATGGAGGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 547
QY 377 cgagccgcgcgtgtgtcgc 436
Db 546 ACAGCAGGGCGCTGTGCTCGCCCCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 487
QY 437 agcaggacgtatgcaagacactgtgcacgcggcagccattcagcctggccgggagccct 496
Db 486 ATCAGCCCGAGTGTCTTGAATAAATGCGTCTCCAAAGGTCCTGACGGGGCTAAGAACGCA 427
QY 497 gccgtcctacagccacgtgtgtcctcaaggaggtgcgctcttcttaacctgcaggtgtct 556
Db 426 GCGACACCTACAGCCACGTTGGTGTCTGTAATCCGTGCGCTTCTTTCAGCTGGAGTCCCTGT 367

BASE COUNT 70 a 177 c 170 g 100 t
ORIGIN

Query Match 23.1%; Score 271.2; DB 11; Length 517;
Best Local Similarity 82.1%; Pred. No. 2.9e-41;
Matches 325; Conservative 0; Mismatches 68; Indels 3; Gaps 1;

```
QY 1 atgtggtgcacaggtttctccagcaagacagtgacagtgcctcctcctgcacagacc 60
Db 125 ATGTGGCTGGCGCGGTCTCCAGCAGCGGGTGACTGCGTCTCTGCTGGCGCAGACCG-- 182
QY 61 tgcctcctgtcttcatactcccgccagggccclcatcccccagccggcgaggat 120
Db 183 -GCCTCCTGTCTTCTCTGATCTCCCGGCCAGACTGCGGCTGCGGCCAGCAGAGGAG 241
QY 121 cgtgtgcagtgctggtgctcctgtggcgctggcgctcctcctcttgggcccagctc 180
Db 242 CGGGTGCATGTGCTAGTGTCTGCTGTGGCGCTCGGCTCGTCTTCTGCTGGCCAGCTC 301
QY 181 ttcagccagcaccccgacgtctcttaoctgatggagccgctggccatgtgtggaccac 240
Db 302 TTCAGCCACACCCCGATGTCTTACCTGATGGAGCGCGGCTGGCACGTGTGGCGCGCC 361
QY 241 ctgtgcagggcagcgcgcaacgctgcacatggcgctgcgcgacctgatgcgtctata 300
Db 362 CTGTGCGAGGGCAGCGCCTTGGCGCTGCACATGGCAGTGGCGACCTGGTGGCTCAGTC 421
QY 301 ttttgtgcacatggacgtgttttgatgcctacatgcacagagccgaaacctgtccgcc 360
Db 422 TTCCGTGTGCACATGATGTGTTCGAGGCCCTACCTGCGCTGGAGGGCGCAACCTGTCCG 481
QY 361 ttttcaactgggcaacgagcgcgcgctgtgctcg 396
Db 482 CTCTTCCAAATGGCGGAGAGCGCGGCTCTGTGCTCG 517
```

```
RESULT 11
AZ405100/c
LOCUS 1M0173108R Mouse 10kb plasmid UUGCLM library Mus musculus genomic
DEFINITION clone UUGCLM0173108 R, DNA sequence.
ACCESSION AZ405100
VERSION AZ405100.1 GI:10529113
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 433)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0173 row: 1 column: 08
Seq primer: CACACAGGAACACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 433.
Location/Qualifiers
FEATURES
source I. .433
/organism="Mus musculus"
```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCLM0173108"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
{http://www.jax.org/resources/documents/dnares/}. The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 113 a 124 c 137 g 59 t
ORIGIN

Query Match 20.5%; Score 240.2; DB 13; Length 433;
Best Local Similarity 75.5%; Pred. No. 1.8e-35;
Matches 312; Conservative 0; Mismatches 98; Indels 3; Gaps 1;

```
QY 1 atgtggtgcacaggtttctccagcaagacagtgacagtgcctcctcctgcacagacc 60
Db 410 ATGCGGCTACCCGGTTCTCCAGCACTGTCATGCTTTCGCTCTGTGTTACAG---ACT 354
QY 61 tgcctcctgtcttctaatctatctcccgccagggccctcctcctccagccggcgaggat 120
Db 353 GGCATCCTGGTCTTCTCTGGTCTCCCGGCAAGTGCCATTGTCCCCAGCAGGCTTGGGGAG 294
QY 121 cgtgtgcacgtgctgctgctcctcctcctcctcctcctcctcctcctcctcctcctc 180
Db 293 CGTGTGCACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 234
QY 181 ttcagccagcaccccgacgtcttcttaoctgatggagcccgctggtgacctgtgtg 240
Db 233 TTCAGCCACACCCCGAIGTTTTCCTGATGGAGCGCGCTTGGCACGTCTGGGATACG 174
QY 241 ctgtgcagggcagcgcgcaacgctgcacatggcgtgcgcgacctgatgcgctctatc 300
Db 173 TTGTGCGAGGGCAGTCCCCCGCACCTCCACATGGCCGTGCGTGACCTGATCCGCTCAGTG 114
QY 301 ttttgtgcacatggacgtgttttgatgcctacatgccacagagccgaaacctgtccgcc 360
Db 113 TTCTTATCCGACATGGACGTATTGATGCTTACCTACCTGCCCTGGCGCCGCAACATCTCGGAT 54
QY 361 ttttcaactgggcaacgagccgcgctgtgctcgcgcgcgcgcgcgcgcgcgcgcgcgc 413
Db 53 CTCTTCCAGTGGGGGTGAGCCGCGCATTTGCTGCTCACCCTCCGCTCTCGGAAGC 1
```

RESULT 12
BG964671

LOCUS BG964671 695 bp mRNA EST 12-JUN-2001
DEFINITION 602831875F1 NCI_CGAP_CO24 Mus musculus cDNA clone IMAGE:4986258 5',
mRNA sequence.

ACCESSION BG964671
VERSION BG964671.1 GI:14352308
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;


```
Db 490 GCCACGAGATCAGCTCACCCCAAGCACTGCAAGCTGCTGTGCGGTACGACGCCCTTTGAT 549
QY 481 ctggcccgaggcctgcgctcctacagccacgctgtgtctcaa-ggaggtgcg-cttct 538
Db 550 ATGGTGGAGAGGCGCTCTCTCAAGGCTCTCGTGTACTCAACGGAGTGGCTATTC 609
QY 539 tcaacctgcaggtgtctatcccgctgtctcagcagaccgcgcgtcacaacctg-cgcatcgtg 597
Db 610 TCAGCCTGCAGGCGCTCTATCCACTGCTCACGGACCCCTTCCCTCAACCTGCCACGTCGTG 669
QY 598 caactggtgcgcgaccgcggcgccgtgtcgtcgtcgtcccgaggag 639
Db 670 CACCTGGTCCGAGAGCCC-CGGGCGCGGTTCGGATCCCGGGAG 710

RESULT 14
BG966340 852 bp mRNA EST 12-JUN-2001
LOCUS 602832826F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4987327 5',
DEFINITION mRNA sequence.
ACCESSION BG966340
VERSION BG966340.1 GI:14353977
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 852)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1098 row: c column: 08
High quality sequence stop: 784.
FEATURES
source
Location/Qualifiers
1..852
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4987327"
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 176 a 248 c 244 g 184 t
ORIGIN

Query Match 18.7%; Score 218.8; DB 11; Length 852;
Best Local Similarity 67.6%; Pred. No. 1.9e-31;
Matches 354; Conservative 0; Mismatches 162; Indels 8; Gaps 3;

QY 124 gtgcagctgctgtgtcctcgtggcgctgggctcctctt-tgggccagctctt 182
Db 260 GTGCAATGTGCTGTCTCTCCTGGCGGTCAGGATCCTCTATATGTGGACAGCTTT 319
QY 183 cagccagcaccgcagctcttctacctgatgagcccgctggtgacatgtgtggaccacct 242
Db 320 CGGGCAGCACCCTGGATGTGTCTACCTGATGAGAGCCCTGCCCTGGGATGTGTGGATGACTTT 379
QY 243 gtgcagggcagcgcggcagcgtgcacatggtgcgacgtgcgcacgtgatgcgtctatctt 302
Db 380 CACCAGCAGCACAGCCTGGAGACTGCACATGGGTGTGCGGGATCTCTGCGTCCGCTT 439
```

```
QY 303 ttgtgcacatggagcgtgtttgatgcctacatg---ccacagagccgaaacctgtccgc 359
Db 440 CCTGTGTGACATGAGCGTCTTTTGTATGCCTACATGAACCCAGGCCCCCGGAACAGTCCAG 499
QY 360 cttttcaactgggcaacgagccgcgcgtgtgtcgcgcgcgcgcgcgcgcgcgccttcc 419
Db 500 CCTCTTCCAGTGGGAGCAAAGCCGCGCCTGTGCTCAGCGCCTGTGTGCTGCTCTTCCC 559
QY 420 ccgaggcaccatcagcaagcaggacgtatgcaagacactgtgcacgcggcagccattcag 479
Db 560 TGCCACAGATCAGCTCACCCCAAGCACTGCAAGCTGCTCTGCGGTACGACGCCCTTTGA 619
QY 480 cctggcccgaggcctgcgcctcctacagccacacctgtgtgtcctcaaggaggtgctctt 539
Db 620 TATGGTGGAGAGGCGCTGCCGATCTCACGGCTTCGTGTGTACTCAAGGAGTGGTATCT 679
QY 540 caacctgcaggtgtctacccgcgtcctcagcagcccgctcacaacctgcgcacctgt--- 596
Db 680 CAGCCTGCAGGCCCTCTATACACTGCTCACGGACCCCTATCCCTCAACCTGGCAACGTCGT 739
QY 597 -gcacctggtgcgcgacccgcggcgctcgtcgtcgtcgtcgtcccgaggag 639
Db 740 GCACCCCTGGTCCGAGACCCCGCGGCGCGGATCCGATCCCGGGAG 783

RESULT 15
BF579746 965 bp mRNA EST 12-DEC-2000
LOCUS 602095056F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4209636 5',
DEFINITION mRNA sequence.
ACCESSION BF579746
VERSION BF579746.1 GI:11653458
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 965)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9775 row: o column: 13
High quality sequence stop: 663.
FEATURES
source
Location/Qualifiers
1..965
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4209636"
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 233 a 300 c 243 g 189 t
ORIGIN

Query Match 13.6%; Score 159.4; DB 11; Length 965;
Best Local Similarity 61.3%; Pred. No. 2.4e-20;
Matches 328; Conservative 0; Mismatches 196; Indels 11; Gaps 4;
```

QY 492 ggcctgcgcgtcctcaagccacgtggtgctcaaggagggtgcgttcttcaacctgcaggt 551
Db 2 GGCGTGCCGCTCTCACGSGCTTCGTGGTACUCAAAGGAGGTGCGTTTCTCAGCCTGCAGGC 61

QY 552 gctctacccgcgtgctcagcgaacccgcgcgtcaacctgcgcacatcgtaacctgggtgcgga 611
Db 62 CCTCTATCCACTGCTCACGGACCCCTTCCCTCAACCTGCACGTGCTGACCTGGTCCGAGA 121

QY 612 cccgcgggcgcgtgctgcgtctcccgaggagcgcgggcccgatactggcaacgcgacaacgg 671
Db 122 CCCCCGGGCGGTGTTCCGATCCCGGGAGCACACCACCATAGAACTCGTGGTTGACAGTCA 181

QY 672 catcgtgctgggcaccacacgggcaa-gtgggtggaggccgacccctcaacctgcgcctgattc 730
Db 182 TATGTCCTAGGGCAGCATTTGGAACGATCAAGGAGGAAGACCCCTATTATGCCATGA 241

QY 731 gcgaggtgtgcgcgcagccacgtgcgcacatgcgcaggccgcacacactcaagccgccacct 790
Db 242 AGATCATCTGCAAAAGCCAGGTGGACATAGTCAGGCCCATCCAAACCCCTCCCTGAAGC-- 299

QY 791 tccgtgcgcgcgcgtacgcgcctggtgcgcttcgaggacctggcggggagcgctggcag 850
Db 300 -TCTGCAGCAGCGCTACCTGTCCTGAGGTATGAGGACCTGGTTCGGGCACCCCTGGCCC 358

QY 851 agatccgcgcactctacgccttcacgcgcctgcgcctcacgcgcacagctcgagccctgga 910
Db 359 AGACGACCAGACTATATAAATT-GTGGGGTTGGATTTTTCGCCCCACCTCCAAACATGGG 417

QY 911 tccacacatcacccacgggtcggggatcggaagccaatcgaggccttccataacttct 970
Db 418 TTCACAAATGTCACCCGGGCAAGGGCATGGGTCAGC-----ATGCCCTCCATACTAAG 471

QY 971 ctagggaatgcgcgaacgtctccacggcctggcgccacgcgttgcccttcaactaa 1025
Db 472 CCAGGAACGCCCTCAACGTCCTCTCAGGCGTGGCGTGGTCTTACCTTACCGAAA 526

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 16:44:36 ; Search time 3220.32 Seconds
(without alignments)
6820.556 Million cell updates/sec

Title: US-09-593-828-3
Perfect score: 2044
Sequence: 1 ggctcgaggtccactgtgct.....aaaaaaaaaaaaaaaaaaaaa 2044

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estfun: *
2: em_esthum: *
3: em_estin: *
4: em_estom: *
5: em_estpl: *
6: em_estba: *
7: em_estro: *
8: em_estov: *
9: em_htc: *
10: gb_est1: *
11: gb_est2: *
12: gb_htc: *
13: gb_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_inv: *
17: em_gss_pln: *
18: em_gss_pro: *
19: em_gss_rod: *
20: em_gss_vrt: *
21: em_gss_other: *

pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	657.4	32.2	695	11	BE857538 7g01a08.x
c 2	631	30.9	735	11	BE858652 7g01a09.x
c 3	525	25.7	525	11	BF197521 7o84a08.x
c 4	487.6	23.9	620	10	AI824100 wj46c01.x
c 5	420.8	20.6	436	10	AW081348 xc41b06.x
c 6	352.4	17.2	1923	12	AK009113 Mus muscu
c 7	346.4	16.9	1067	13	CNS03KN7
c 8	308.4	15.1	517	11	BF042384
c 9	295.4	14.5	568	10	AI115260
c 10	279.4	13.7	849	13	CNS04QEN
c 11	259.4	12.7	955	11	BGI07354
c 12	245.4	12.0	433	13	AZ405100

c 13	231.4	11.3	337	11	BG960153
14	224.4	11.0	695	11	BG964671
15	224.4	11.0	783	11	BG963298
16	218.8	10.7	852	11	BG966340
17	217.4	10.6	624	13	AQ373399
c 18	213.8	10.5	560	10	AL589348
19	211.4	10.3	452	11	BF725761
20	211	10.3	352	10	AW089016
21	210.2	10.3	400	10	AV760391
22	210	10.3	388	13	AQ557029
23	209.2	10.2	412	10	BE062478
24	209	10.2	539	13	AQ379787
25	208.4	10.2	1172	12	AF305824
c 26	208.2	10.2	416	10	AW341978
27	207.8	10.2	616	13	AQ554309
c 28	207.2	10.1	451	10	AV695478
c 29	207.2	10.1	650	13	AQ530872
c 30	207	10.1	384	10	AL040054
31	207	10.1	946	11	BG335756
32	206.6	10.1	602	13	AQ540860
c 33	206.4	10.1	345	11	BG222875
34	206	10.1	544	11	BG739841
35	206	10.1	544	13	AQ394650
36	205.6	10.1	440	11	BG189911
c 37	205.4	10.0	400	10	AV760389
c 38	205	10.0	550	10	AL138431
39	205	10.0	599	10	AW970932
40	204.8	10.0	695	11	BG430795
41	204.6	10.0	672	13	AQ631465
c 42	203.8	10.0	327	10	AA483606
43	203.8	10.0	415	13	AQ150411
44	203.8	10.0	481	10	AW964231
c 45	203.6	10.0	495	10	AW835445

ALIGNMENTS

RESULT 1
BE857538/c
LOCUS
DEFINITION
7g01a08.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3305174 3'
similar to TR:Q9UBY3 Q9UBY3 N-ACETYLGLUCOSAMINE
6-O-SULFOTRANSFERASE. ; mRNA sequence.
BE857538
BE857538.1 GI:10371664
EST
EST
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 695)
AUTHORS
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE
National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTCAP), Tumor Gene Index
JOURNAL
Unpublished (1998)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 431.
Location/Qualifiers
1. 695
/organism="Homo sapiens"

LOCUS BE858652 735 bp mRNA EST 29-SEP-2000
DEFINITION 7901a09.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3305176 3'
similar to TR:Q9UBY3 Q9UBY3 N-ACETYLGUCOSAMINE
6-O-SULFOTRANSFERASE. ; mRNA sequence.
ACCESSION BE858652
VERSION BE858652.1 GI:10373890
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 735)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
Unpublished (1998)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 425.
FEATURES
Location/Qualifiers
1..735
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3305176"
/clone_lib="NCI_CGAP_Brn23"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGCGCATATCTTTTCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT 135 a 221 c 253 g 120 t 6 others
ORIGIN
Query Match 30.9%; Score 631; DB 11; Length 735;
Best Local Similarity 93.5%; Pred. No. 2.6e-63;
Matches 677; Conservative 0; Mismatches 45; Indels 2; Gaps 2;
QY 300 cccggccaggccctcatccccagccggcgaggagatcgtgtgcacgtgctgt 359
Db 734 CCCGCNcAGGGCCCTCATCCACAAGCAGCAGNCGCGAAGATCTGTGCACGTGTGGTGTGT 675
QY 360 cctcgtggcgtggggtcattccttcttggggcagctcttcagccagcaccgccgacct 419
Db 674 CCTCGTGGCGCTCGGCTCATCTCTNAGGGCCAGCTCTCAGCCAGCA-CCCGACGTCT 616
QY 420 tctacctgatggagcccgctggtggcatgtgtggaccacctgtcgcagggcagcgcgcaa 479
Db 615 TCTACATGATGAAGCCCGCGTGGCATGTGTGG-CCACCCCTGTGCGAGGGCAGCGCGCAA 557
QY 480 cgctgcacatggccgtgcggacctgatgcgctctatcttttgcgacatggacgtgt 539
Db 556 CGCTGCACATGGCCGTGCGCGACCTGATGCGCTCTATCTTTTGTGCGACATGGACGTGT 497
QY 540 ttgatgcctacatgccacagagcgaaacctgtccgccttttcaactgggcaacgagcc 599

/db_xref="taxon:9606"
/clone="IMAGE:3305174"
/clone_lib="NCI_CGAP_Brn23"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGCGCATATCTTTTCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT 127 a 210 c 247 g 111 t
ORIGIN
Query Match 32.2%; Score 657.4; DB 11; Length 695;
Best Local Similarity 97.6%; Pred. No. 2.6e-66;
Matches 678; Conservative 0; Mismatches 16; Indels 1; Gaps 1;
QY 284 ctgctcttcattcatctccggccaggccctcatccccagccggcgaggagatcgtgtg 343
Db 695 CTGCTTTCATCATCTGCCGCGCAGGGCCATCATCACAGCGCGCAGCCGAGCATCGTGTG 636
QY 344 cacgtgctggtgctgctcctcgtggcgctcgtggtcctccttcttgggcccagctcttcagc 403
Db 635 CACGTGCTGATGCTGTCTCTGCTGGCGTCCGGCTCATCC-TCCTGGGACAGCTCTTCAGC 577
QY 404 cagcaccgccagctcttctacctgatggagcccgctggtgcatgtgtggaccacctgtcg 463
Db 576 CAGCACACCGACGCTCTTCTACTTGATGGAGCCCCGCGTGGCATGTGTGGACCATCTGTG 517
QY 464 cagggcagcgcggaacgctgcacatggccgctgcgacacctgatgcgctctatctttttg 523
Db 516 CAGGGCAGCGCGGCAACGCTGCACATGGCCGTGGCGGACGATGCGCTCTATCTTTTG 457
QY 524 tgcgacatggacgtgtttgatgcctacatgccacagagccgaaacctgtccgcttttcc 583
Db 456 TGGGACATGCACGTGTCTGATGCCTACATGCCACAGAGCCGAAACCTGTCCGCTATTTC 397
QY 584 aactgggcaacgagccgctgctgtgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 643
Db 396 AACTGGGCAACGAGCGCGCGCTGTGCTGCCGCCGCGCTGCAGCGCCCTTCCCGGAGGC 337
QY 644 accatcagcaagcaggacgtatgcaagacactgtgcagcggcagccattcagccctggcc 703
Db 336 ACCATCAGCAAGCAGGACGATGCAAGACACTGTGCACGCGGCAGCCATTACGCCCTGGCC 277
QY 704 cgggaggcctgcgctcctacagccacgtggtgctcaaggaggtgcgcttcttcaacctg 763
Db 276 CCGGAGGCCCTGCCGCTCCTACAGCCACGTGTGTGCTCAAGGAGTGCCTTCTTCAACCTG 217
QY 764 cagtgctctaccgctgctcagcgacccccgcgctcaacctgcgcacatcgtgcacctggtg 823
Db 216 CAGGTGCTCTACCCGCTGCTCAGCGACCCCGCGCTCAACCTGCGCATCGTGCACCTGGTG 157
QY 824 cgggacccgcgccgctgctgcctcccgaggcgcgcgccgagccgatactggcagcgac 883
Db 156 CCGGACCCGCGGGCCCGTGTGCTGCCCTCCCGGAGGCGCGCGCCCGGATCTGGCACGCGAC 97
QY 884 aacggcatcgtgctgggcaaccaacggcaagtgggtggaggccgacctcacctgcgcctg 943
Db 96 AACGGCATCGTGTGCTGGGCAACCAACGGCAAGTGGGTGGAGCCCGACCCCTCACCTGCGCCTG 37
QY 944 attcgcgaggtgtgcccagccacgtgcgcatcgc 978
Db 36 ATTCCGAGGTTGTGCGCGCAGCCACGTCGTCATCGC 2

TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:969488
Seq primer: custom primer used
high quality sequence stop: 517.

```

FEATURES
source
Location/Qualifiers
1. .568
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1885164"
/clone_lib="Sugano mouse embryo mewa"
/dev_stage="embryo, 14 dpc"
/lab_host="DH10B"
/note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[TTGTTGGCCTACTGG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
CACCATGTG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science).
Custom primers for sequencing: 5' end primer
CTTCCTGCTCTAAAGACTGCG and 3' end primer
CCATGCTCCAGCTTCAGCCACA."

```

BASE COUNT	80 a	189 c	173 g	126 t
ORIGIN				
Query Match		14.5%;	Score 295.4;	DB 10;
Best Local Similarity		75.4%;	Pred. NO. 3.9e-25;	Length 568;
			Micro-matches 121.	Indels 3;
				Gaps 1;

QY	202	gccccagcgccgcgatgtggctgccacggttctccagcaagacagtacagtgctcct	261
Db	67	GGCCCTAGCCAGAGGTAIGCGGTACCCCGTTTCTCCAGCACTGTATGCTTTTCGCTCCT	126
QY	262	cctggcacagaccacctgcctoctgctctttcatcatctcccggccaggccctcatcccc	321
Db	127	GATGGTACAG--ACTGGCATCTGGTCTTCTCCTGGTCTCCCGGAAGTGCCATCGTCCCC	183
QY	322	agccggcgcgaggatcgtgtgcacgtgctgggtgctgctctctggtgcgctcgggctcatc	381
Db	184	AGCAGGCCCTTGGGAGCGGTGCACGTGCTGGTACTGTCTCTCGTGGCGCTCGGGCTCGTC	243
QY	382	cttcttggccagctcttcagccagcaccgccgacgtcttctaacctgatggagcccgctg	441
Db	244	CMTCTGGGCCAGCTCTTCAGCCAACACCCCGATGCTTCTACCTGATGAGACCGGCTTG	303
QY	442	gcattgtlggaccacctgtgcgagggcagcgcggcaacgctgcacatggccgtggcgga	501
Db	304	GCACGCTGGGATACGTTGTGCGAGGGCAGTGCCCCCGCACTCCACATGGCCGTCGGTGA	363
QY	502	cctgatgcgctctatctttttgtgcacatggacagtgtttgatgcctacatgccacagag	561
Db	364	CCTGATCCGCTCAGTGTTCTCTAATGCGACATGGACGTATTGTATGCCCTACCTGCCCTGGCG	423
QY	562	ccgaaacctgtccgccttttttaactggccaagagccgcgcgctgtgctcgcgcgccgc	621
Db	424	CCGAAACATCTCGGATCTCTTCCAGTGGGCGGTGACCGCGCATIGTGCTCACCTCCGGT	483

D _b	484	CTGCGAAGCCTTCGCTCTGGCAACATCAGCAGCGAGGAGGTGTCTAAGCCTCTGTGCGC	543
Q _Y	682	gcggcgagcattcagcctggcccg	706
D _b	544	AACGCGGCCCTTCGGCCTGGCTCAG	568

RESULT	10
CNS04QFN/c	
LOCUS	
DEFINITION	CNS04QFN 849 bp DNA GSS 24-MAY-2000 Tetraodon nigroviridis genome survey sequence pUC-Ori end of clone 129C06 of library G from Tetraodon nigroviridis, genomic survey sequence.

AL302540
AL302540.1 GI:8181872
GSS: genome survey sequence.
Tetraodon nigroviridis.
Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

REFERENCE
AUTHORS

1 (bases 1 to 849)
Roest-Crolius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.

TITLE	Charaterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
COMMENTS	Unpublished

JOURNAL REFERENCE AUTHORS	Unpublished 2 (bases 1 to 849)	
	Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.	
TITLE	Human gene number estimate provided by genome wide analysis using retroceder niaroviridis DNA sequence	

JOURNAL
REFERENCE
AUTHORS
Unpublished
3 (bases 1 to 849)
Genoscope.

Accession	Direct Submission	EMBL/GenBank/DBJ databases
U00001.1	Submitted (12-APR-2000)	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .

FEATURES source

```
1. .849
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="129006"
/clone_lib="G"
/note="Genoscope sequence ID : C0BG129BH03SP1-end :
pTC-ori"
```

BASE COUNT	137 a	248 c	273 g	184 t
ORIGIN				/ other

Query Match	13.7%	Score 279.4;	DB 13;	Length 849;
Best Local Similarity	60.8%	Pred. No. 2e-23;		
Matches 483; Conservative		3; Mismatches 305;	Indels 3;	Gaps 2;

QY 602 gccgtgtgcgcgcccgcctgcagcgccctttccccaggaccattcagcaagcaggac 661
| | | | | | | | | | | | | | | | | | | | | |
+
603 gggatggaagg 765

Db
824 GGGCTGTGCTGGCCCCCGCGCTGTTCTCTCACCCCCCGGAGGATGAGCGATCAGCCC

[illegible]

Db 764 GAGTGCTTGAAAAATCGGTGCGCAAGGGGTCTGCAAGGGGCTACGAGGCGCTACCGCGT 781

Qy 722 tacagccacgtggtgctcaaggaggtgcgcttctttcaacctgcagggtgctctaccgctg 781

704 TACAGCCACGTTGGTGCTGAAATCCGTGCGCTTCTTCGAGTGGAGTCCCTGTATCCGCTC 645
 Db

TITLE and Wright,D.,Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0173 row: I column: 08
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 433.

FEATURES
source

1. .433
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0173I08"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT
ORIGIN

113 a 124 c 137 g 59 t
Query Match 12.0%; Score 245.4; DB 13; Length 433;
Best Local Similarity 74.7%; Pred. No. 2.2e-19;
Matches 322; Conservative 0; Mismatches 106; Indels 3; Gaps 1;
QY 200 tggcccccagccgcgcagtggtggcgcacggttctccagcaagacagtgacagtgctc 259
Db 428 TAGGCCCTAGCCAGAGGTATGCGGTACCCGGTTTCTCCAGCACTGTCTATGCTTCCGCTC 369
QY 260 ctccctggcacagaccacccctcctcctcttcatcatctccggccagggccctcatcc 319
Db 368 CTGATGGTACAG---ACTGGCATCCTGGTCTTCTCCTGGTCTCCGGCAAGTGCCATTGTC 312
QY 320 ccagccggcgccgaggatcgtgtgcacgtgctggtgctgctcgtcgtggcgtcgggctca 379
Db 311 CCAGCAGGCCTTGGGGAGCGTGTGCACGTGCTGGTACTGTCTCGTCTCGTGGCGCTCG 252
QY 380 tccctttggccagctcttccagccagcaccgccgacgtcttctacctgatggagcccgcg 439
Db 251 TCCTTTCGTGGGCCAGCTTTTCAGCCCAACACCCCGATGTTTTCCTGATGGAGCCGCT 192
QY 440 tggcatgtgtggaccaccctcgtgcagggcagcgcggaacagctgcacatggccgtgcgc 499
Db 191 TGGCACGCTCTGGGATACGTTGTTCGAGGCGAGTGCCCCCGCACCTCCACATGGCCGTGCGT 132

QY 500 gacctgatcgctctatctctttttgtggacatggacggtgtttgatgcctacatgccacag 559
Db 131 GACCTGATCCGCTCAGTGTTCCTATGCGACATGGACGATTTTGATGCTTACCTGCCCTGG 72
QY 560 agccgaaacctgtccgcctttttcaactgggcaacgagccgcgcgtgtgctgcgcgcc 619
Db 71 CGCCGCAACATCTCGGATCTCTTCCAGTGGCGGTGAGCCGGCATTTGCTACCTCCG 12
QY 620 gcctgcagcgc 630
Db 11 GTCTGCGAAGC 1

RESULT 13
BG960153/c

LOCUS BG960153 337 bp mRNA EST 12-JUN-2001
DEFINITION PM3-CT0640-300301-003-cl2 CT0640 Homo sapiens cDNA, mRNA sequence.
ACCESSION BG960153
VERSION BG960153.1 GI:14378324
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 337)

REFERENCE

AUTHORS
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663

JOURNAL

MEDLINE
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil

Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM3&t2=PM3-CT0640-
300301-003-cl2&t3=2001-03-30&t4=1)

Seq primer: puc 18 forward
High quality sequence start: 20
High quality sequence stop: 126.

FEATURES

Location/Qualifiers
1..337
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0640"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT
ORIGIN

96 a 84 c 87 g 70 t
Query Match 11.3%; Score 231.4; DB 11; Length 337;
Best Local Similarity 91.0%; Pred. No. 9.8e-18;
Matches 292; Conservative 0; Mismatches 21; Indels 8; Gaps 4;

QY 1502 ggagtttggggtccctccctgaagtaagcaaggactgcacgtttctctctctgattc 1561
|||||


```
Db 337 GGAGTTTGGGGTCTCCCTCGAAGTAGGCAAGGACTGCACGGTTTCTTTCTCTCTGATC 278
QY 1562 tcggttttcctttgagtccttctgagctgccttc-----tcatcaggtgcactcttcag 1616
Db 277 TCGGTTTCTCTTGGTCTTCTGAGCTCTTCTGAGCTGCCCTTCGTCATGCAAGGTGCACCTCTTTCATIG 218
QY 1617 gaaaagcaactc-ttgcctctactctcttctggcgagggagtaagttactgc-taaatt 1674
Db 217 GAAATGCAACTCTTTGCCCTACCTCTTCTGGCGCAGGAGTAAGTTACTGCTTAAATT 158
QY 1675 aaattaaatgtgtccagggccgggtgctggtggctcactgctgtaataccagcattttgag 1734
Db 157 AAATTAAATGTGTCCAGCCAGGCTACAGTGGCTCATGCTGTAATCCAGCATTTTGAG 98
QY 1735 aggtgagggcgggtggatcacctgaggtcagga-ttcaaaaccagcctggccaaacatagt 1793
Db 97 AGGCTGAGGCGGTGGATCACCTGAGTTCAGGAGTTCGAAACCAAGCCTGGCCCAACATAGT 38
QY 1794 gaaacccccctctactactaaa 1814
Db 37 GATACCCCTCTCTACTAACA 17

RESULT 14
BG964671 695 bp mRNA EST 12-JUN-2001
LOCUS 602831875F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4986258 5',
DEFINITION mRNA sequence.
ACCESSION BG964671
VERSION BG964671.1 GI:14352308
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 695)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM1095 row: f column: 19
High quality sequence stop: 684.
Location/Qualifiers
BASE COUNT 118 a 216 c 195 g 166 t
ORIGIN
Query Match 11.0%; Score 224.4; DB 11; Length 695;
Best Local Similarity 65.6%; Pred. No. 4e-17;
Matches 391; Conservative 0; Mismatches 196; Indels 9; Gaps 4;

QY 284 ctgctcttcattcatctccggccagggccctcatccccagccggg---cgaggatcg 339
Db 99 CTCCTCATCCATATGTCCGTGCCACAGACACCTTCCAGAGGGGAGGAGTCCAGGAGGCC 158
```

```
QY 340 tgtgcacgtgctgtgctgctctctctgctgctggctcggctcctctcttctggccagctctt 399
Db 159 CGTGCAATGTGCTGGTGTCTTCTCCTGGCGGTTCAGGATCCT-CTTTGTGGGACAGCTTTT 217
QY 400 cagccagcaaccccgacgtcttctacctgatggagcccgctggcatgtgtggaccaccc 459
Db 218 CGGGCAGCACCCCGGATGTGTTCTACCTGATGGAGCCTGCCCTGGCATGTGTGGATGACTTT 277
QY 460 gtcgcagggcagcgcggaacgctgcacatggccgtgctgcgcgacctgatgcgctctatctt 519
Db 278 CACCAGCAGCACAGCCCTGGAAGCTGCACATGGCTGTGCGGATCTTCTGCGTTCCGTTCTT 337
QY 520 tttgtgcacatggacgtgtttgatgcctacatg---ccacagagccgaacccgtgcgc 576
Db 338 CCTGTGTGACATGAGCGTCTTTGATGCTTACATGAACCCAGGCCCGCCGGAACAGTCCAG 397
QY 577 ctttttcaactgggcaacgagcccgctgctgctgcgcgcgcgcgcgcgcgcgcgcgcgc 636
Db 398 CCTCTCCAGTGGGAGCAAAAGCCGGGCCCTGTGCTCAGCGCCTGTGTGTGACTTCTTCC 457
QY 637 ccgagggcaccatcagcaagcaggacgtatgaagacacactgtgcacggcgccacattcag 696
Db 458 TGCCACAGGATCAGCTCACCACCAAGCAGCTGCAGAGCTGCTCTGCGGTGAGCAGCCCTTGA 517
QY 697 cctggcccgaggcctgctcctcactacagccacgctgtgtcctcaaggaggtgcgctctt 756
Db 518 TATGTGGAGAAAGCCCTGCGCTCTCAGCGCTCGTGTACTCAAGGAGGTGCGTTTCT 577
QY 757 caactgcaggtgctctaccgctgctcagcagccccc-cgcgctcaacctgcgcctctgc 815
Db 578 CAGCCTGCAGGCCCTCTATCCACTGCTCAGCGACCCCTTCCCTCAAGCTGCAGCTCCGTG 637
QY 816 acctgtgcgcgaccccgcgccgtgctgcctccccgggagggcgccggcccgata 871
Db 638 CACCTGTCCGAGACCCCGGCGCGTGTCCGATCCCGGGGAGCCACACCACTTA 693

RESULT 15
BG963298 783 bp mRNA EST 12-JUN-2001
LOCUS 602827716F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4982643 5',
DEFINITION mRNA sequence.
ACCESSION BG963298
VERSION BG963298.1 GI:14350935
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 783)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM1095 row: p column: 04
High quality sequence stop: 741.
Location/Qualifiers
BASE COUNT 118 a 216 c 195 g 166 t
ORIGIN
Query Match 11.0%; Score 224.4; DB 11; Length 695;
Best Local Similarity 65.6%; Pred. No. 4e-17;
Matches 391; Conservative 0; Mismatches 196; Indels 9; Gaps 4;

QY 284 ctgctcttcattcatctccggccagggccctcatccccagccggg---cgaggatcg 339
Db 99 CTCCTCATCCATATGTCCGTGCCACAGACACCTTCCAGAGGGGAGGAGTCCAGGAGGCC 158
```

Fri Feb 1 20:20:54 2002

/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. Constructed by Life
Technologies.. Note: this is a NCI_CGAP Library."

BASE COUNT 141 a 236 c 226 g 180 t
ORIGIN

| | | | | |
|-----------------------|---|--------------------|-----------------|-------------------|
| Query Match | 11.0%; | Score 224.4; | DB 11; | Length 783; |
| Best Local Similarity | 70.7%; | Pred. No. 3.8e-17; | | |
| Matches 369; | Conservative | 0; | Mismatches 146; | Indels 7; Gaps 5; |
| QY 341 | gtgcacgtgctggtgctgcctcctgctcgtgctgggctcattccttcttggtccagctcttc | 400 | | |
| Db 190 | gtgcattgtgctggtgctgcttccctggcgtcaggatccctcttttgggacacagcttttc | 249 | | |
| QY 401 | agccagcaccccgacgtctttacacctgtagggcccgctggcattgtgtggaccacctg | 460 | | |
| Db 250 | ggcagcaccccgatgtgttacctgtagggccctggcattgtgtggatgactttc | 309 | | |
| QY 461 | tcgcaggcgagcgggcaacgctgcacatggccgtgctgagacctgatgcgtcttatctt | 520 | | |
| Db 310 | accagcacacagcctgggaagctgcacatggctgctgggattcttctgggttccgtcttc | 369 | | |
| QY 521 | ttgtgcacatggacgtgtttgatgcctacatg---ccacagagccgaacctgtccgcc | 577 | | |
| Db 370 | ctgtgtgacatgagcgtctttgatgcctacatgaaccagccagcccccggaaacagtcacc | 429 | | |
| QY 578 | tttttcaactgggcaacagccgcgcgtgtgctgcgcgcgcgcctgcagcgcccttccc | 637 | | |
| Db 430 | ctcttccagtgaggagcaaaagccggccctgtgctcagcgcctgtgtgacttcttccct | 489 | | |
| QY 638 | cgaggcaccatcagcaagcaggacgtatgcaagacactgtgcacgcggcagccattcagc | 697 | | |
| Db 490 | gcccacgacatcagctcaccacagcactgcaagctgctctgctgctgctgctgctgctg | 549 | | |
| QY 698 | ctggcccgaggagcctgcccctcctacagccacgtggtgctcaa-ggaggtgcg-cttct | 755 | | |
| Db 550 | atggtggagaaaggcctgcccctctcaccggcttcggtggtactcaacggaggtgcgtattcc | 609 | | |
| QY 756 | tcaacctgcaggtgctctacccgctgctcagcagcccccgcgtcaacctg-cgcattcgtg | 814 | | |
| Db 610 | tcagcctgcagccctctctatccactgctcagggaccttccctcaacctgccacgctcgig | 669 | | |
| QY 815 | cacctggtgcgcgaccccgggcgctgctgcgtcccgaggag | 856 | | |
| Db 670 | cacctgggtccgagaccc-cgggcccgtgttcggatccccgggag | 710 | | |

Search completed: January 31, 2002, 16:44:48
Job time: 5772 sec